OPTIMAL ROAD PRICING SCHEME DESIGN

AGACHAI SUMALEE

Submitted in accordance with the requirements for the degree of
DOCTOR OF PHILOSOPHY

The University of Leeds
Institute for Transport Studies

August 2004

To my late sister, Wanida Sumalee (Orn)
"The flower of my heart"

Acknowledgements

I would like to sincerely thank my supervisors, Prof. Tony May and Dr. Simon Shepherd for their patience, support, and advice throughout this research, especially when I was in the hardest time of my life. The first time I learnt about road pricing was in Tony's lecture when I was a Master student, and from that moment I started getting really interested in this research topic. Thanks for such a good lecture, Tony. Simon, I think we have come a long way together. Thanks for being such a good supervisor, colleague, and friend.

My life and work in Leeds has been very enjoyable. I have to thank Prof. David Watling (aka Dave) for being such a good colleague, boss, friend, and travel companion. Thanks a lot, Dave. I am sure we will carry on our in-car discussion about various transport research topics for a very long time (well, also about Leeds United and Liverpool FC).

Also, I have to thank my colleagues and friends at ITS and, in particular, the Network Group for supporting me in various ways throughout the last three and half years (some names to mention in particular: Dave Milne and Richard Connors).

Special thanks go to Prof. Siriphong Lawphongpanich for his time and discussion on the cutting plane algorithm and optimal toll problem, to Prof. Hai Yang for suggesting a number of useful comments that help improving my thesis, to Ferri Ghahri-Saremi for his initial suggestion on the branch-tree idea, to Nick Marler for being such a nice office mate and your help on my writing.

I'd like to acknowledge the financial support from the UK Department for Transport that made this research possible. Also, I want to thank Mike Goodwin and Phil Philippou who were looking after the students funded by the DfT for the last four years. I would also like to thank all participants of the survey conducted during this research.

Furthermore, I would like to thank Mr. David Kamnitzer at IBI group (London) for his support at the initial stage of this study and made this study possible.

I am deeply indebted to my fiancée, Aya Fukushige, who is always there for me and understand me throughout the long hard period of my study. Thanks for your companion, patience and love. Also, my sincere gratitude goes to Aya's parents who have been very supportive for both of us.

My personal gratitude goes to my father, mother, brother, and sister for their moral support and encouragement. My parents always give the priority to education which really had a good influence on us and made it possible for me to pursue my PhD study.

Life is so uncertain. I'd like to dedicate this thesis to my beloved sister, Wanida Sumalee (Orn), who passed away at such a young age. You are always proud of me and love me with all your heart. I am also proud of you and love you so much. Sleep well my sister. You will always be with me and our family forever.

"We must be willing to let go of the life we have planned, so as to have the life that is waiting for us"

"The three great essentials to achieve anything worthwhile are first, hard work; second, stick-to-it-iveness; and third, common sense." Thomas Edison

Abstract

There are two main approaches to designing road pricing schemes. The first is judgmental in nature and focuses on the acceptability and practicality of the scheme. The second is based on theory concentrating on the optimality and performance of the scheme. This research aimed to integrate these two approaches into a single framework and to develop a tool to aid the decision maker in designing a practical and optimal road pricing scheme.

A review of the practical design criteria and a survey with six local authorities in the U.K. were conducted to clarify the concept of the judgmental design. A simple charging scheme like a charging cordon is believed to be the most practical charging regime due to its simple structure. The decision on the boundary and structure of the cordon is based largely on public acceptability and possible adverse impacts. Road pricing is used to serve several objectives including congestion reduction, revenue generation, and increase in efficiency of the transport system.

The framework for the theoretical optimal toll design problem adopted was a Stackelberg game where the travellers' behaviour were assumed to follow the concept of Wardrop's user equilibrium. This problem can also be formed as a Mathematical Program with Equilibrium Constraint (MPEC). After reviewing various methods for solving the MPEC problem, three possible methods (the merit function method, improved cutting plane algorithm, and Genetics Algorithm (GA) based algorithm) were developed and tested with the optimal toll problem. The GA based algorithm was found to be the most appropriate for the development of the design algorithm with practical constraints.

Three different features of the judgmental design were included into the optimisation algorithm: the closed cordon formation, constraints on the outcomes of the scheme, and the allowance for multiple objectives. An algorithm was developed to find the optimal cordon with an optimal uniform toll. It is also capable of designing a scheme with multiple cordons. The algorithms for solving the constrained optimal cordon design problem and the multiobjective cordon design problem were also developed. The algorithm developed for the multiobjective problem allows the application of the posterior and progressive preference articulation approach by generating the set of non-dominated solutions.

The algorithms were tested with a network of Edinburgh. The results revealed several policy implications. Adopting a judgmental cordon with a simple uniform toll may be less effective. A variable optimised toll around the judgmental cordon can generate around 70% more benefit than the optimal uniform toll. The optimised location of a cordon generated about 80% higher benefit compared to the best judgmental cordon. Additional constraints such as a maximum of total travel time decreased the level of the benefit of the scheme by 90%. Different objectives may require different designs for the charging cordon scheme. The welfare maximisation cordon should focus on those trips contributing most to the social welfare function which are mainly in the congested areas with an appropriate toll level. The revenue maximisation cordon should impose a higher number of crossing points and minimise possible diversion routes to avoid the tolls which should be high. The equity cordon should cover a wider area of the network with low toll level to ensure a good distribution of the cost and benefit to all origin-destination pairs.

The algorithms developed can offer support to the decision maker in developing a charging cordon scheme by formalising the process of charging cordon design. This will increase the transferability of the technique and the transparency of the decision process.

Table of Content

ACKNOWLEDGEMENTS I

ABSTRACT III

TABLE OF CONTENT V

LIST OF TABLES IX

LIST OF FIGURES XI

CHAPTER 1 INTRODUCTION 1

- 1.1 TRANSPORT PROBLEMS AND EMERGENCE OF ROAD PRICING 1
- 1.2 APPROACH TO DESIGNING A ROAD PRICING SCHEME 5
  - *1.2.1 Judgmental approach* *5*
  - *1.2.2 Theoretical approach* *7*
- 1.3 OBJECTIVES AND SCOPE OF THE RESEARCH 9
- 1.4 STRUCTURE OF THE THESIS 10

CHAPTER 2 ROAD PRICING: FROM THEORY TO PRACTICE 14

- 2.1 INTRODUCTION 14
- 2.2 DEVELOPMENT OF THE ECONOMIC THEORY 15
- 2.3 ROAD PRICING AS A TRANSPORT INSTRUMENT 21
  - *2.3.1 Objective of the road pricing scheme* *21*
  - *2.3.2 Possible methods for implementation* *23*
- 2.4 EXPERIENCES AROUND THE WORLD 24
  - *2.4.1 Developments in the U.K.* *24*
  - *2.4.2 Developments in other countries in Europe* *27*
  - *2.4.3 Developments in Singapore* *33*
  - *2.4.4 Development of ERP in Hong Kong* *36*
  - *2.4.5 Developments elsewhere in the world* *38*
- 2.5 RESEARCH ISSUES 40
  - *2.5.1 Public and political acceptability* *40*
  - *2.5.2 Equity impact and social exclusion* *43*
  - *2.5.3 Economic impact* *45*
  - *2.5.4 Technological aspect of the implementation* *46*
- 2.6 SUMMARY 48

CHAPTER 3 JUDGMENTAL APPROACH 50

- 3.1 INTRODUCTION 50
- 3.2 REVIEW OF THE PRACTICAL DESIGN CRITERIA 50
  - *3.2.1 Design options* *52*
  - *3.2.2 Dealing with the objectives of the scheme* *53*
  - *3.2.3 Dealing with public acceptability* *54*
  - *3.2.4 Dealing with adverse impacts* *56*
- 3.3 ISSUES ON DESIGN CRITERIA AND SURVEY DESIGN 58
- 3.4 SURVEY RESULTS 60
  - *3.4.1 Responses to the questionnaires* *60*
  - *3.4.2 In-depth interview analysis* *65*
- 3.5 FINDING FROM THE SURVEY 69
- 3.6 SUMMARY 71

CHAPTER 4 FORMULATION OF OPTIMAL TOLL DESIGN PROBLEM 73

- 4.1 INTRODUCTION 73
- 4.2 USER'S EQUILIBRIUM CONDITION 74
  - *4.2.1 Notations* *74*
  - *4.2.2 Different formulations of UE condition* *76*
- 4.3 MPEC FORMULATION OF THE OPTIMAL TOLL PROBLEM 85
  - *4.3.1 MPEC and BLPP framework* *85*
  - *4.3.2 Variations of optimal road pricing design problems* *88*

4.4 POSSIBLE OUTCOME INDICATORS 91
4.5 SUMMARY 100

CHAPTER 5 OPTIMISATION ALGORITHM FOR OPTIMAL TOLL PROBLEM 102

5.1 INTRODUCTION 102
5.2 DEVELOPMENT OF ALGORITHMS IN THE TRANSPORT AREA 104
  5.2.1 Iterative optimisation approach 105
  5.2.2 Linearisation approach 106
  5.2.3 Sensitivity analysis 107
  5.2.4 Inverse optimisation 108
  5.2.5 Minmax/maxmin based optimisation 109
  5.2.6 Smoothing approach and reduced UE condition 110
  5.2.7 Cutting plane algorithm 112
  5.2.8 Meta-heuristic optimisation approach 112
5.3 A TAXONOMY OF MPEC ALGORITHMS: OPTIMISATION PERSPECTIVE 113
  5.3.1 Algorithms for the linear and quadratic BLPP 114
  5.3.2 Algorithms for the general BLPP/MPEC 116
5.4 SMOOTHING ALGORITHM: MERIT FUNCTION APPROACH 124
  5.4.1 Problem with the complementarity constraint 124
  5.4.2 Smoothing the complementarity constraint 129
  5.4.3 Tests of the merit-function approach with small network 132
5.5 IMPROVED CUTTING PLANE ALGORITHM APPROACH 136
  5.5.1 Cutting plane algorithm for the second-best toll problem 136
  5.5.2 Tests of the improved cutting plane algorithm with Pacman network 142
5.6 GENETIC ALGORITHM APPROACH 150
  5.6.1 Genetic algorithms 150
  5.6.2 GA-CHARGE: the method for optimising the toll level 151
  5.6.3 Experiments with GA-CHARGE 154
5.7 SUMMARY 161

CHAPTER 6 OPTIMAL CHARGING CORDON DESIGN 164

6.1 INTRODUCTION 164
6.2 OPTIMAL CHARGING CORDON DESIGN PROBLEM 166
  6.2.1 Formulating the road pricing design problem 166
6.3 BRANCH-TREE FRAMEWORK 168
  6.3.1 Definition of a charging cordon 168
  6.3.2 Nomenclature of the branch-tree concept 171
  6.3.3 Relationship between branch-tree and closed charging cordon 174
  6.3.4 Dummy nodes in a branch-tree 176
  6.3.5 Two-way link issue 178
6.4 APPLYING GA TO SOLVE THE OPTIMAL CLOSED CORDON 179
  6.4.1 Chromosome design 181
  6.4.2 Initialisation 183
  6.4.3 Evaluation process 184
  6.4.4 Selection process 185
  6.4.5 Crossover process 186
  6.4.6 Mutation process 188
6.5 GA-AS FOR MULTIPLE CHARGING CORDON DESIGN 189
  6.5.1 Required condition for multi-layer charging cordons 190
  6.5.2 Hyper-graph approach to define multiple charging cordons 191
  6.5.3 GA implementation 192
6.6 TESTS WITH THE EDINBURGH NETWORK 193
  6.6.1 Description of the tests 193
  6.6.2 Comparison between GA-AS with grid-search and toll-vector 194
  6.6.3 Optimal cordon without a constraint on toll level 198
  6.6.4 Optimal cordon for each toll level 202
  6.6.5 Multiple cordons design 203
  6.6.6 Overall comparison 205
  6.6.7 Analysing the effect of GA parameters 206
6.7 SUMMARY 211

CHAPTER 7 OPTIMAL CHARGING CORDON DESIGN WITH CONSTRAINTS 215
7.1 INTRODUCTION 215
7.2 GA APPROACHES FOR CONSTRAINED OPTIMISATION PROBLEMS 216
  7.2.1 Constraint relaxation method 217
  7.2.2 Restricting search inside the feasible region 220
  7.2.3 Hybrid methods 221
7.3 CONSTRAINED CHARGING CORDON DESIGN: SELF-ADAPTIVE PENALTY METHOD 222
7.4 TESTS WITH THE EDINBURGH NETWORK 223
  7.4.1 Comparison between static, dynamic, and self-adaptive penalty methods 224
  7.4.2 Constrained charging cordon design 227
7.5 SUMMARY 235

CHAPTER 8 CHARGING CORDON DESIGN WITH MULTIPLE OBJECTIVES 237
8.1 INTRODUCTION 237
8.2 DECISION APPROACH TO MULTIOBJECTIVE PROBLEM 238
  8.2.1 Preference articulation 239
  8.2.2 Decision paradigms for the multi-objective problem 241
8.3 REVIEW OF GA APPROACHES FOR MULTIOBJECTIVE OPTIMISATION PROBLEM 243
  8.3.1 Classical methods 243
  8.3.2 Evolutionary based methods 246
8.4 NSGA-II FOR OPTIMAL CORDON DESIGN: POSTERIOR PREFERENCE ARTICULATION APPROACH 250
  8.4.1 Overview of the method 250
  8.4.2 Crowding distance assignment: Crowding-sort $(F_l, <_c)$ *253*
  8.4.3 Crowding tournament selection operator 254
  8.4.4 Clustering method for representing the Pareto solutions 255
8.5 NSGA-II WITH PROGRESSIVE PREFERENCE ARTICULATION 255
  8.5.1 Revisit the definition of non-dominated solution 255
  8.5.2 Progressive preference articulation for NSGA-II 260
8.6 TESTS OF NSGA-II WITH EDINBURGH NETWORK 263
  8.6.1 Priori preference articulation case 264
  8.6.2 Posterior preference articulation case 268
  8.6.3 Progressive preference articulation case 284
8.7 SUMMARY 288

CHAPTER 9 POLICY IMPLICATIONS 291
9.1 INTRODUCTION 291
9.2 PERFORMANCE COMPARISON 291
  9.2.1 Effect of toll levels and locations 291
  9.2.2 Trade-off between practical constraints and benefits 302
  9.2.3 Compromised cordon design with multiple objectives 307
9.3 POSSIBLE APPLICATIONS AND BENEFITS OF THE APPROACH TO REAL WORLD DECISIONS 311
9.4 SUMMARY 315

CHAPTER 10 CONCLUSIONS AND FURTHER RESEARCH 318
10.1 THE GAP BETWEEN PRACTICAL AND OPTIMAL ROAD PRICING DESIGN 318
10.2 JUDGMENTAL DESIGN APPROACH 319
10.3 THEORETICAL DESIGN APPROACH 320
10.4 INTEGRATION OF THE JUDGMENTAL DESIGN AND THEORETICAL DESIGN 323
10.5 SOME POLICY IMPLICATIONS 326
10.6 RECOMMENDATIONS FOR FURTHER RESEARCH 327

BIBLIOGRAPHY 333

APPENDIX A. QUESTIONAIRE ON JUDGMENTAL DESIGN CRITERIA FOR CHARGING CORDON 352

APPENDIX B. BACKGROUND OF THE CASE STUDY 354

APPENDIX C. MATHEMATICAL NOTATIONS 357

APPENDIX D. THE LINK COST AND OD DEMAND FUNCTIONS FOR THE EDINBURGH NETWORK 358

List of Tables

Table 2-1 Key characteristics of the Norwegian toll rings (Source: May and Sumalee, 2003) 28

Table 3-1 Possible design criteria for practical a road pricing scheme design based on literature review 58

Table 3-2 Response to question regarding the objective of the scheme 61

Table 3-3 Responses to questionnaires regarding the design issue related to adverse impacts of the scheme 63

Table 3-4 Responses to questionnaires regarding the design issue related to practicality of the scheme 64

Table 3-5 Responses to questionnaires regarding the design issue related to public acceptability of the scheme 64

Table 5-1 Test results with single tolled link with the five-link network (smoothing approach) 134

Table 5-2 Iteration results for tolling on link 5 of the five-link network (smoothing approach) 135

Table 5-3 Test results with dual tolled links with the five-link network (smoothing approach) 135

Table 5-4 Link travel time function parameters for Pacman network 143

Table 5-5 Parameters of the demand functions for Pacman network 143

Table 5-6 Optimal single toll results with the Pacman network 144

Table 5-7 Test results with two and three tolled links (Pacman network) 148

Table 5-8 Iterates from the improved CPA with test 4 150

Table 5-9 Parameters used for GA-CHARGE with Pacman test 154

Table 5-10 GA-CHARGE results with Pacman network 155

Table 5-11 GA-CHARGE parameters for the tests with the Edinburgh network.... 158

Table 5-12 Results from the tests of GA-CHARGE with the Edinburgh network.. 158

Table 6-1 Parameters used for GA-AS 195

Table 6-2 Summary of the results of designing a single optimal charging cordon for Edinburgh network 206

Table 7-1 Levels of violations and penalty coefficients for static penalty function 224

Table 7-2 Performance indicators for the do-nothing scenario and the tests with judgmental cordons 228

Table 7-3 Overall results for constrained cordon designs..229

Table 8-1 Results of the weighted sum approach (priori preference articulation) for MOOP-CD 265

Table 8-2 Non-dominated solutions from the MOOP, maximising net benefit and net revenue..271

Table 8-3 Non-dominated solutions from the MOOP, maximising net benefit and minimising gini coefficient..274

Table 8-4 Non-dominated solutions from the MOOP, maximising net benefit and minimising total travel time..276

Table 8-5 Clusters of the non-dominated solutions (maximizie net benefit and minimise total travel distance)..278

Table 8-6 Clusters of the non-dominated solutions (maximizie net benefit and net revenue and minimise Gini coefficient)...283

Table 9-1 Marginal cost tolls from system optimum and GA-CHARGE tolls for the top-15 links of the Edinburgh network 293

Table 9-2 Comparison of the performance of different charging regimes for the Edinburgh network...293

Table 9-3 Statistics on travel time/trip, travel distance/trip, and average speed for different charging scheme of the Edinburgh network......................................300

Table 9-4 Comparison of the scheme benefits with additional constraints..............303

List of Figures

Figure 1-1 Average traffic speeds in Greater London: 1980/82- 2000/03 (Source: Office for National Statistics, 2003) 1
Figure 1-2 Structure of the thesis 11
Figure 2-1 Illustration of the economic theory of marginal cost pricing 15
Figure 4-1 Users -Planner interaction 86
Figure 4-2 Graphical illustration of the definition of economic benefit 93
Figure 4-3 Graphical illustration of the definition of economic benefit improvement 94
Figure 4-4 Graphical illustration of the definition of economic benefit improvement based on rule of half 95
Figure 4-5 Lorentz curve explaining the distributional impact 97
Figure 5-1 Milestone of the development of the transport NDP 104
Figure 5-2 Feasible region generated by the complementarity condition 126
Figure 5-3 Illustrative example of the linear dependent problem with MPEC 128
Figure 5-4 Multiplier sets of the illustrative example 128
Figure 5-5 Contours of the smoothed F-B function (with $\rho = 10$) 130
Figure 5-6 Contours of the min function 130
Figure 5-7 Five-link network used for testing the smoothing algorithm for solving COTP 133
Figure 5-8 Revenue curve as the function of the toll on link 1 (five-link network) 134
Figure 5-9 Structure of Pacman network for the tests with the improved CPA 143
Figure 5-10 Iteration of the improved CPA method optimising the toll on link 2-4 (Pacman) 145
Figure 5-11Iteration of the improved CPA method optimising the toll on link 7-6 146
Figure 5-12 Examples of the iterations of the improved CPA (link 1, 3, 4, and 13)147
Figure 5-13 Contour plot of the objective function and iterations of the improved CPA (link 17 and link 13) 148
Figure 5-14 Objective function surface and iterations of the improved CPA (link 17 and link 1) 149
Figure 5-15 Convergence of the Master objective and True objective functions with CPA 149
Figure 5-16 Overall framework of GA-CHARGE 152

Figure 5-17 Chromosome structure for GA-CHARGE 153
Figure 5-18 Crossover process in GA-CHARGE 154
Figure 5-19 Chromosomes generated by GA-CHARGE for test no. 1 with Pacman network 156
Figure 5-20 Best solution found so far and average fitness in each generation from the test of GA-CHARGE with Pacman network (test no. 4) 156
Figure 5-21 Edinburgh network with three judgmental cordons 157
Figure 5-22 Best solution found so far and average fitness in each generation from the test of GA-CHARGE with the Edinburgh network (inner1 cordon)......... 159
Figure 5-23 Best solution found so far and average fitness in each generation from the test of GA-CHARGE with the Edinburgh network (outer cordon)........... 159
Figure 5-24 Optimal tolls for the inner1 cordon found by GA-CHARGE (highlighted links are tolled links)... 160
Figure 5-25 Optimal tolls for the outer cordon found by GA-CHARGE (highlighted links are tolled links)... 160
Figure 6-1 Example of drawback of cutset definition.. 170
Figure 6-2 Example of branch-tree and branching process..................................... 173
Figure 6-3 Example of sub-branch.. 173
Figure 6-4 Relationship between branch-trees and charging cordons 176
Figure 6-5 Dummy node in branch-tree... 177
Figure 6-6 Two-way link issue for branch-tree formulation................................... 179
Figure 6-7 Overall process of GA-AS.. 180
Figure 6-8 Chromosome structure for GA-AS... 181
Figure 6-9 Chromosome structure for toll vector.. 185
Figure 6-10 Crossover process for branch-tree... 187
Figure 6-11 Mutation process for branch-tree... 189
Figure 6-12 Definition of multiple charging cordons scheme 190
Figure 6-13 Defining multiple charging cordons with cuts 191
Figure 6-14 Chromosome structure for multiple charging cordons......................... 192
Figure 6-15 Edinburgh network and three judgmental charging cordons................ 195
Figure 6-16 Performance comparison between GA-AS with grid-search and toll-vector strategies.. 196
Figure 6-17 Fitness distribution of GA-AS with toll-vector.................................... 197
Figure 6-18 Fitness distribution of GA-AS with grid-search.................................. 197

Figure 6-19 Optimal cordon solution found by GA-AS 198

Figure 6-20 Variation of total benefits and revenues for OPC1 cordon with different toll levels 199

Figure 6-21 Cordons with 2nd and 3rd highest net total benefits 200

Figure 6-22 Location of OPC-REV (maximise net revenues) 201

Figure 6-23 Variation of the benefit from optimal charging cordon for different fixed toll levels 202

Figure 6-24 Locations of optimal cordons for toll levels of £0.50, £0.75, £1.25, £2.00, £3.00 203

Figure 6-25 Optimal double charging cordons for the Edinburgh network 204

Figure 6-26 Variation of the net benefit of the optimal double cordon with different toll levels 205

Figure 6-27 Comparison of the performance of GA-AS with different values of probability of crossover (Pc) 208

Figure 6-28 Comparison of the performance of GA-AS with different values of probability of mutation (Pm) 208

Figure 6-29 Profile of the improvement of the best solution from the GA-AS with the different values of Pc 209

Figure 6-30 Profile of the improvement of the best solution from the GA-AS with the different values of Pm 209

Figure 6-31 Moving averages per 25 generations of the variance of the fitness values within each generation for different values of probability of mutation 210

Figure 7-1 Average net benefits and net revenues within each generation (static penalty method) 226

Figure 7-2 Average net benefits and net revenues within each generation (dynamic penalty method) 226

Figure 7-3 Average net benefits and net revenues (self-adaptive penalty) 227

Figure 7-4 Optimal cordon location with constraint on net revenues (CON-REV) with toll level of £2.00 230

Figure 7-5 Optimal cordon location with constraint on equity impact (CON-GINI) with toll level of £0.75 231

Figure 7-6 Users' benefits distribution for different charging cordons 231

Figure 7-7 User benefits by O-D pairs for OPC1 232

Figure 7-8 User benefits by O-D pairs for CON-REV 232

Figure 7-9 User benefits by O-D pairs for CON-GINI 232
Figure 7-10 Optimal cordon location with constraint on total travel distance (CON-TTDIS) with toll level of £2.00 233
Figure 7-11 Optimal cordon location with constraints on equity impact and net revenues (CON-GINI-REV) 234
Figure 8-1 Non-dominating level 251
Figure 8-2 Schematic of the NSGA-II algorithm 253
Figure 8-3 Illustration of the crowding distance 254
Figure 8-4 Framework for the progressive preference articulation method with NSGA-II 260
Figure 8-5 Illustration of focusing effect on the search as the results from decision maker's goals set 263
Figure 8-6 Results for MOOP-CD with weighted sum approach (highlighted links are toll links) 266
Figure 8-7 Elements of the best solution found so far for the weighted sum problem (WSum1) 267
Figure 8-8 Elements of the best solution found so far for the weighted sum problem (WSum2) 267
Figure 8-9 Pareto solutions and corresponding Pareto front for MOOP problem with objectives of maximising net benefits and net revenues 270
Figure 8-10 Weigthed sum objectives with different weight pairs (weight for net benefit, weight for net revenue) 272
Figure 8-11 Pareto solutions and corresponding Pareto front for MOOP problem with objectives of maximising net benefits and minimising *Gini* coefficient (equity impact) 273
Figure 8-12 Location of WG1 cordon (with toll level of £0.50) 273
Figure 8-13 Location of WG31 cordon (with toll level of £1.50) 273
Figure 8-14 Pareto solutions and corresponding Pareto front for MOOP problem with objectives of maximising net benefits and minimising total travel time 276
Figure 8-15 WT1 Cordon with the optimal toll level of £4 277
Figure 8-16 Pareto solutions and corresponding Pareto front for MOOP problem with objectives of maximising net benefits and minimising total travel distance ... 278
Figure 8-17 Relationship between the total travel distances and total demand in the network 279

Figure 8-18 WD6 and WD7 cordons 280
Figure 8-19 Comparison of link flows between WD6 and WD7 280
Figure 8-20 Pareto surface and non-dominated solutions for the MOOP-CD with three objectives (net benefit, net revenue, and equity) 281
Figure 8-21 Scatter plots for the MOOP-CD results with three objectives (net benefit, net revenue, and Gini coefficient) 282
Figure 8-22 Pareto surface and 15 clusters of the non-dominated solutions for the MOOP-CD with three objectives (net benefit, net revenue, and equity) 283
Figure 8-23 Pareto front and progressive goals set after 40 generations of NSGA-II 284
Figure 8-24 Pareto front and progressive goals set after 80 generations of NSGA-II 285
Figure 8-25 Pareto front after 120 generations of NSGA-II 286
Figure 8-26 Pareto front and progressive goals set after 160 generations of NSGA-II 286
Figure 8-27 Pareto front from the progressive preference articulation after 200 generations 287
Figure 9-1 Top-15 links with the highest marginal social cost toll for the Edinburgh network (highlighted link with the label indicating the rank) 293
Figure 9-2 Link travel delay in the do-nothing scenario of the Edinburgh network (the width of the bandwidth represents the level of delay on that link) 294
Figure 9-3 Comparison of the link delay for the outer cordon with the uniform toll and variable toll (uniform – variable) 296
Figure 9-4 Plot of the trips passing through links A and B in the do-nothing scenario 296
Figure 9-5 Comparison of the link flows from the tests with OPC1 and inner2 cordons 298
Figure 9-6 Comparison of the link flows from the Top-15 and OPC1 charging scheme 299
Figure 9-7 Comparison of link flows with the inner2 and outer cordons 300
Figure 9-8 Current proposed double charging cordon for the Edinburgh scheme (Source: http://www.tiedinburgh.co.uk/congest.html) 304
Figure 9-9 Variation of total benefits and revenues for OPC1 cordon with different toll levels 305

Figure 9-10 Comparison of the link flows from OPC1 and CON-GINI....................306
Figure 9-11 Location of OPC-REV as found by GA-AS for maximising the net revenue (highlighted links are tolled links)...308
Figure 9-12 Location of the WG1 cordon (highlighted links are tolled links)........310
Figure 9-13 O-D benefits with the WG1 cordon (£0.50 toll)...................................310
Figure 9-14 O-D benefits with the OPC1 cordon (£1.50 toll)..................................311

CHAPTER 1 INTRODUCTION

All learning begins with the simplest phrase, "I don't know".

1.1 TRANSPORT PROBLEMS AND EMERGENCE OF ROAD PRICING

If we ask someone from any major urban city the question "what is the major problem in your city?", surely one of the most frequent answers we could hear is *'traffic congestion'*. Is this true? Recent statistics from the U.K. Department for Transport (DfT) reports that average traffic speeds in Greater London between year 2000-2003, before the implementation of the congestion charging scheme, were lower than in 1980 (Office for National Statistics, 2003), see Figure 1-1 . This problem is also reflected in the recent U.K. National Census Survey with over half of the respondents reported to consider the congestion problem in towns and cities as a serious issue for them (Office for National Statistics, 2001). The recent survey commissioned by the U.K. DfT also reported that congestion is ranked as the most serious transport problem (Department for Transport, 2001).

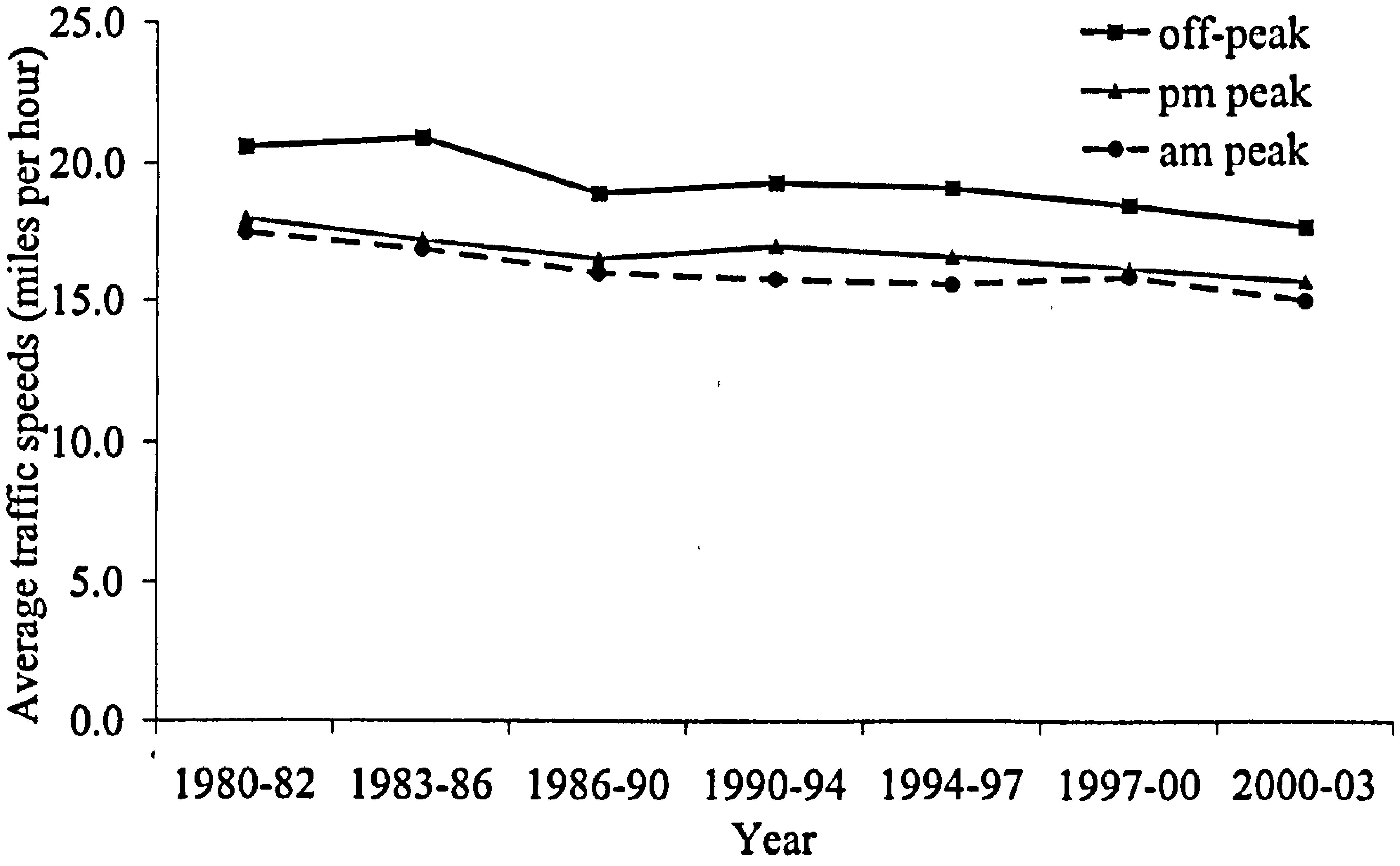


Figure 1-1 Average traffic speeds in Greater London: 1980/82- 2000/03 (Source: Office for National Statistics, 2003)

The main impact of congestion is the delay imposed on the travellers and the inefficiency imposed on the transport and economic system. Overall, congestion costs in OECD countries are estimated to be equivalent to around 2 percent of GDP (OECD, 1995). In the U.S., the cost of delay in urban areas was estimated to be about $43 billion in 1990 (Shrank *et al*, 1993). Similarly in the U.K., traffic congestion is claimed to be costing the U.K industry around £15 billion per year (Confederation of British Industry, 1994).

The impact of congestion is not only the delay and inefficiency. It also comes with the major environmental impact. In the U.K., it was estimated that road traffic is responsible for at least 90%, 20% and 49% of the overall levels of CO, CO_2 (the principal greenhouse gas), and NO_x gas respectively (Department of Environment/Scottish Office, 1996).

This increasing congestion problem of the induced traffic demand (SACTRA, 1994) has encouraged the central governments in various countries to consider a new direction for transport policy. The idea of traffic restraint or traffic demand management (TDM) has emerged as a possible method to effectively control the travel demand. To this end, the idea of road pricing has re-emerged as a possible method to restrain the travel demand. In the U.K., the government has recently acknowledged the necessity of road pricing as the key transport policy by legislating the introduction of road user charging:

> *"Many of our towns and cities face significant levels of congestion and pollution .... But experience has shown that improving public transport and related traffic management measures whilst necessary are not sufficient in many cases. We will therefore introduce legislation to allow local authorities to charge road users so as to reduce congestion, as part of a package of measures in a local transport plan that would include improving public transport"* (Ministry of Transport, 1998 page 103)

This change in the legislation and the realisation of the need for a pricing measure to curb the congestion problem resulted in the implementation of the congestion charging scheme in Durham in 2002 and in London in 2003. The U.K. government also set up the Congestion Charging Development Partnership to support the development of charging schemes in U.K. cities. Current major developments of the charging schemes include the proposals in Edinburgh and Bristol.

Apart from the development of the road pricing policy in the U.K., major steps were also made in other parts of the world. Singapore continues to be the city with the most advanced development of the application of road pricing in their integrated transport policy. The first Area Licensing Scheme was implemented in Singapore in the 1970s. In Norway, a number of toll ring schemes area under operation in different cities (e.g. Oslo, Trondheim, and Bergen), whose objectives were originally to finance transport infrastructure, are currently in a transition to the congestion charging based system.

The simple idea of road pricing is that by imposing an additional cost for a car trip the travel demand made by car should decrease naturally. Thus, road pricing seems to emerge as an effective way to reduce travel demand in the cities. However, the other justification of road pricing is based on the principle of marginal cost pricing.

From an economic perspective, the idea of road pricing has been proposed for more than seventy years. Economists claimed that road users do not pay the true price of their road use which leads to inefficiency in the transport system (Pigou, 1920; Walters, 1961; Vickrey, 1969). There exist externalities (in the form of congestion, pollution, and accidents imposed by car use) that the car users do not perceive and are not charged for.

With this simple but yet powerful argument, the concept of marginal cost pricing for road users was proposed in which it was suggested that an additional price can be imposed upon the road user (in the form of a road toll) to correct the market failure. If the additional price is set appropriately, the social welfare of the system will be maximised.

This economic theory is the main reason for the inclusion of the pricing policy in the recent EU transport White Paper (Commission of European Community, 2001):

"Part of the reason for this situation (congestion problem) is that transport users do not always cover the costs they generate. Indeed, the price structure generally fails to

reflect all the costs of infrastructure, congestion, environmental damage, and accidents." (Commission of European Community, 2001 page 8)

As described, road pricing originated from the idea of marginal cost pricing to encourage the efficient use of the roads, and in parallel its application has evolved to be one of the most effective traffic demand management measures to curb the congestion problem. However, regardless of the philosophy behind the idea of road pricing the basic concept is deceptively simple; apply the price mechanism to control the road use in the same way as it is applied elsewhere (e.g. telecommunication and electricity). Intellectually, the question is one of determining the appropriate price to ensure the efficient usage of the infrastructure. In other words, the practical questions for the implementation of a road pricing system concern "how to charge", "where to charge", "when to charge", "whom to charge", and "how much to charge". Often, finding out the answers to these questions is not a simple task given the complex nature of the transport system and the responses of the road users to the charge. In addition, road pricing causes a serious public and political acceptability problem. Inevitably, the determination of the charging system should also consider issues related to the acceptability and practicality of the scheme.

On one hand, a road pricing scheme can be designed to achieve the highest benefit possible (depending on the objective of the scheme). On the other hand, the scheme can be designed to ensure its acceptability and practicality. Alternatively, can we design a scheme that balances both needs? This research is indeed at the intersection between these two approaches and aims to answer this question. The incentive of the research is twofold. One is to tackle the challenge of devising an appropriate method to integrate the optimal and practical design of a road pricing scheme. The other is the emerging need for an appropriate tool to aid the design of the road pricing schemes. The remainder of this chapter provides a brief overview of the development of the approach to designing a road pricing scheme, the objectives of this study, the methodology and scope, and finally the structure of the thesis.

1.2 APPROACH TO DESIGNING A ROAD PRICING SCHEME

1.2.1 Judgmental approach

Road pricing is a very politically sensitive policy. Although the initial idea was proposed almost seventy years ago and the first U.K. research programme on this policy was commissioned in the 1960s (Ministry of Transport, 1964), the first U.K. full scale road pricing scheme was only implemented in 2003 in London, forty years after the first study. The key problem has been the public and political acceptability of the scheme. Obviously, charging what used to be free will provoke the opposition of those who are affected by the scheme. The U.K. is not the only country to suffer from delays in implementing a successful road pricing scheme. There were also similar cases around the world in Hong Kong, the Netherlands, and Sweden. The delay in these cases was mainly caused by the problems of public acceptability, political acceptability, and practicality of the schemes. Borins (1988) even suggested that the idea of road pricing may never be practical and implementable due to the political and public acceptability issues.

With the realisation of the possible barrier to the implementation of road pricing, a number of studies have been conducted to analyse the approach and strategy to avoid the public confrontation of the implementation of road pricing (see for example Goodwin, 1989; Giuliano, 1992; Gomez-Ibanez, 1992; Sheldon *et al*, 1993; Dittmar *et al*, 1994; Giuliano, 1994; Rom, 1994; May *et al*, 1996; Langmyhr, 1997; Jones, 1998; Rietveld and Verhoef, 1998; Fridstrom *et al*, 2000; Ison, 2000; Milne *et al*, 2001; Jones, 2002; Jaensirisak *et al*, 2003; Schade and Schlag, 2003). Probably, the most notable study was the Smeed Report (Ministry of Transport, 1964) that suggested a number of key requirements for a road pricing scheme design (mainly involving effectiveness, acceptability and practicality of the scheme).

With the complex political and practical issues involved, it was found that the approach adopted in the design and feasibility study of a road pricing scheme in the real world is mainly based on professional judgment and experience of the designer. The first collection of evidence on the way in which a road pricing scheme was designed was the study of the Area Licensing Scheme (ALS) in Singapore (Holland and Watson, 1978). Some of the key features and criteria considered for the design of the scheme include the consideration of the diversion effect of the traffic, the coverage of the scheme over the present and future congested area of the city, simplicity of the scheme, and the effectiveness of the scheme in reducing congestion in the central business district (Holland and Watson, 1978).

The pilot study for Electronic Road Pricing (ERP) in Hong Kong was a further step forward in the way in which a road pricing scheme is designed (Transpotech, 1985). Different charging cordons and screen lines were determined by quantitatively analysing the congestion level and potential benefits of imposing the tolls on different links. However, the resulting design options involved a number of cordons and screen lines and were considered too complex. The complexity of the scheme required a high technological solution to implement it and this caused a public acceptability problem due to arguments over the invasion of privacy of the scheme (Harrison, 1986).

Little evidence on the actual process and criteria for the design is available. Probably, this is due to the political sensitivity of the policy and the lack of a definite process for the design of a road pricing scheme. In most cases, traffic modelling is used to forecast the responses of the users to the road pricing scheme as the post-evaluation tool for a number of scheme options judgmentally predefined. This process could naturally lose some of the benefit of road pricing through the "trial and error" process of the design (May *et al*, 2002).

Some studies suggested a number of key considerations for ensuring the acceptability and practicality of a road pricing scheme. Sheldon *et al* (1993), based on their survey, suggested that a simple charging scheme would be more acceptable to the public (e.g. a charging cordon scheme). Based on a questionnaire survey Jaensirisak *et al* (2003) established the relationship between the scheme design, the toll level and the acceptability of a road pricing scheme.

1.2.2 Theoretical approach

The initial theoretical development of the idea of road pricing relied on a number of simplistic assumptions (Sharp, 1966). For instance, the original example of the idea of road pricing proposed by Pigou (1920) was largely based on the assumption of the first-best condition of pricing in the economy. The first-best condition is satisfied when the prices of all competing travel alternatives (alternative routes or modes) are set according to their marginal costs. Such an assumption ignores the nature of the complex and imperfect environment in reality.

There may be a high transaction cost of implementation and enforcement which makes the first-best pricing (impose tolls on all links) impossible. The simple introduction of first-best road tolls over a subset of the links of a network may distort the allocation of the traffic over the network and may cause a degradation in the welfare instead of improving the situation. In some cases, the price of competing modes (e.g. bus and train) may not be under the control of the local authority. Thus, the idea of setting the toll equal to the marginal cost of the trips as suggested in the early developments may not be valid for the design of an optimal road pricing scheme in a more complex environment.

The early development of theory for the optimal road pricing problem has been the inclusion of an explicit representation of the network and travellers' responses to the toll into the analysis. The travellers are normally assumed to be rational economic agents who seek to maximise their utility or minimise their dis-utility of travel. The network is represented as a set of connected links and origins and destinations of travel demand. The basic behaviour included in the analysis is the route choice. The concept of Wardrop's user equilibrium (Wardrop, 1952) is normally adopted to govern the route choice behaviour of the travellers. With the Wardrop's equilibrium, the traveller from each origin-destination pair will choose her route so as to minimise her generalised travel cost. The Wardropian equilibrium condition is a special case of Nash's equilibrium (Nash, 1951).

With this simple assumption about the travellers' behaviour in the network, a number of economists have analysed the determination of the optimal road toll with alternative untolled routes in the network. This represents the recognition of the necessity to include the "second-best" nature of the competition (due to the existence of the untolled alternative routes) into the analysis of the optimal toll problem. Levy-Lambert (1968) and Marchand (1968) were probably the first to analyse the optimal toll problem in a two-link network with one untolled link. Later on Verhoef (2002) analysed the optimality condition of the second-best toll problem for a general network and proposed a heuristic algorithm for determining the optimal tolls for a subset of links.

The problem of defining an optimal road pricing charge can be framed as a Stackelberg game (Stackelberg, 1952) that involves the leader and the followers in the game. The followers are assumed to follow Nash's equilibrium. The leader has the ability to anticipate the reaction of the followers to the strategy he/she choose. In the setting of the optimal toll problem, the leader can be considered as the authority responsible for setting the toll for a road pricing scheme. The followers, thus, are the road users whose behaviours are assumed to follow Wardrop's equilibrium. The Stackelberg game can also be considered as a Mathematical Program with Equilibrium Constraint, MPEC, (Luo *et al*, 1996). Finding the solution of the MPEC is a very difficult task due to the complexity of the equilibrium condition imposed as the constraint of the optimisation problem.

If all the links in the network can be tolled, i.e. first-best condition, the problem can be solved as a different optimisation problem which is in a simpler form, minimising the total travel cost (for the fixed demand case) or maximising social welfare (for the elastic demand) (Sheffi, 1985; Yang and Bell, 1997). However, if only a subset of links can be tolled, the optimal toll problem becomes a MPEC problem.

The framework of MPEC and the Stackelberg game are the foundation of the development of the approach to determining the second-best optimal toll with untolled routes. It has attracted interests from other disciplines (e.g. engineering, operations research, and computing). A number of methods have been proposed for solving MPEC which are mostly applicable to the optimal toll problem (see for instance Luo *et al*, 1996; Marcotte and Zhu, 1997; Shimizu *et al*, 1997; Bard, 1998; Outrata *et al*, 1998; Facchinei *et al*, 1999; Marcotte *et al*, 2001; Hearn and Yildirim, 2002).

The idea of the theoretical design of a road pricing scheme concentrates on the optimality and performance of the road pricing scheme. With the framework of MPEC, the problem of road pricing design is transformed to an optimisation problem. The network modelling framework is adopted to represent the system. The aim is then to define the prices/charges on the set of tolled links so that the equilibrium is guided to the state that maximises the objective function set. If the objective function is the social welfare function, then the resulting toll regime is the second-best optimal toll. Obviously, the main drawback of this theoretical approach is the lack of consideration of the practical constraints of the scheme (e.g. public acceptability and closed cordon format).

1.3 OBJECTIVES AND SCOPE OF THE RESEARCH

The research presented here aims to *develop an approach that integrates the optimal design of a road pricing scheme from the economic perspective with the practical requirements of the scheme from the professional point of view.* This integrated approach aims to support the decision making process of the practitioner in developing and designing the road pricing scheme. The decision making process involves the decision on the trade-off between satisfying the design requirements and the overall benefit of the road pricing scheme. It also involves the trade-off between the achievements of the scheme for different sub-objectives. With this overall goal of the research, a number of objectives for the study can be defined as follows:

- to review the current development and practice of the design for a practical road pricing scheme ;
- to investigate and define the requirements for a practical road pricing scheme from the professional and judgmental perspective;
- to investigate the approach to deriving the optimal design of a road pricing scheme from the theoretical point of view;
- to develop a mathematical approach to designing an optimal road pricing scheme taking into account the practical requirements;
- to develop a quantitative approach to support the decision making on the trade-off between the benefit of the scheme and the satisfaction of the design constraints;

- to develop an approach to support the trade-off of different objectives of a road pricing scheme through the design of the scheme;
- to test the algorithms developed with a realistic size network and demonstrate the application of the methods to the real world problem;
- to draw policy conclusions from the modelling results.

The scope of this study is limited mainly by the modelling assumptions. The main model assumption adopted is that of static Wardrop's user equilibrium which assumes users have perfect knowledge. The model also assumes a single user class, a single time period, and a single mode. The suppression of the demand is determined by an elastic demand function. The possible responses of the users to the toll scheme are change of route and the decision to travel or not to travel. There is no land-use response in the model. In addition, the congestion represented in the model is assumed to depend directly and solely on the flow volume on the link. The interaction at junctions is omitted for simplicity. However, it should be noted that the algorithm developed, which will be presented in the thesis, is not limited to these assumptions.

1.4 STRUCTURE OF THE THESIS

The aim in this thesis is to integrate qualitative and quantitative approaches to the development of a road pricing scheme. On one hand, the idea of road pricing stems from rigorous economic theory. On the other hand, the implementation of this policy involves many aspects of politics, urban management, and social issues. It is a very challenging task to include the materials from both perspectives comprehensively in one single document. Each chapter is written such that it can be read independently. A number of examples are provided in each chapter to aid the explanation. Given the variety of the research topics and issues investigated in this thesis, a review of the relevant research is provided in each chapter.

The thesis is structured into four parts (see Figure 1-2). Part I (Chapter 1- Chapter 3) concentrates on the background of road pricing and the overview of the research presented in this thesis. The development of road pricing both from transport policy and economic perspectives is explained in Chapter 2. This chapter gives a clear view of the development of road pricing from economic theory and its transition to a practical transport policy. Various experiences on planning and implementing road pricing schemes around the world are reviewed in this chapter. Some related and future research challenges for improving the understanding of the impact and benefit of road pricing are also discussed in Chapter 2.

Chapter 3 then focuses on the detail of the judgmental design approach for a road pricing system. A review of practical design criteria for a road pricing system is presented accompanied by the results from probing questionnaires and interview surveys with various local authorities in the U.K., who are members of the U.K. congestion charging development partnership. A summary of the criteria adopted in designing a road pricing scheme found from the survey is discussed. These design criteria will then be integrated into the theoretical design of an optimal road pricing scheme later in Part III.

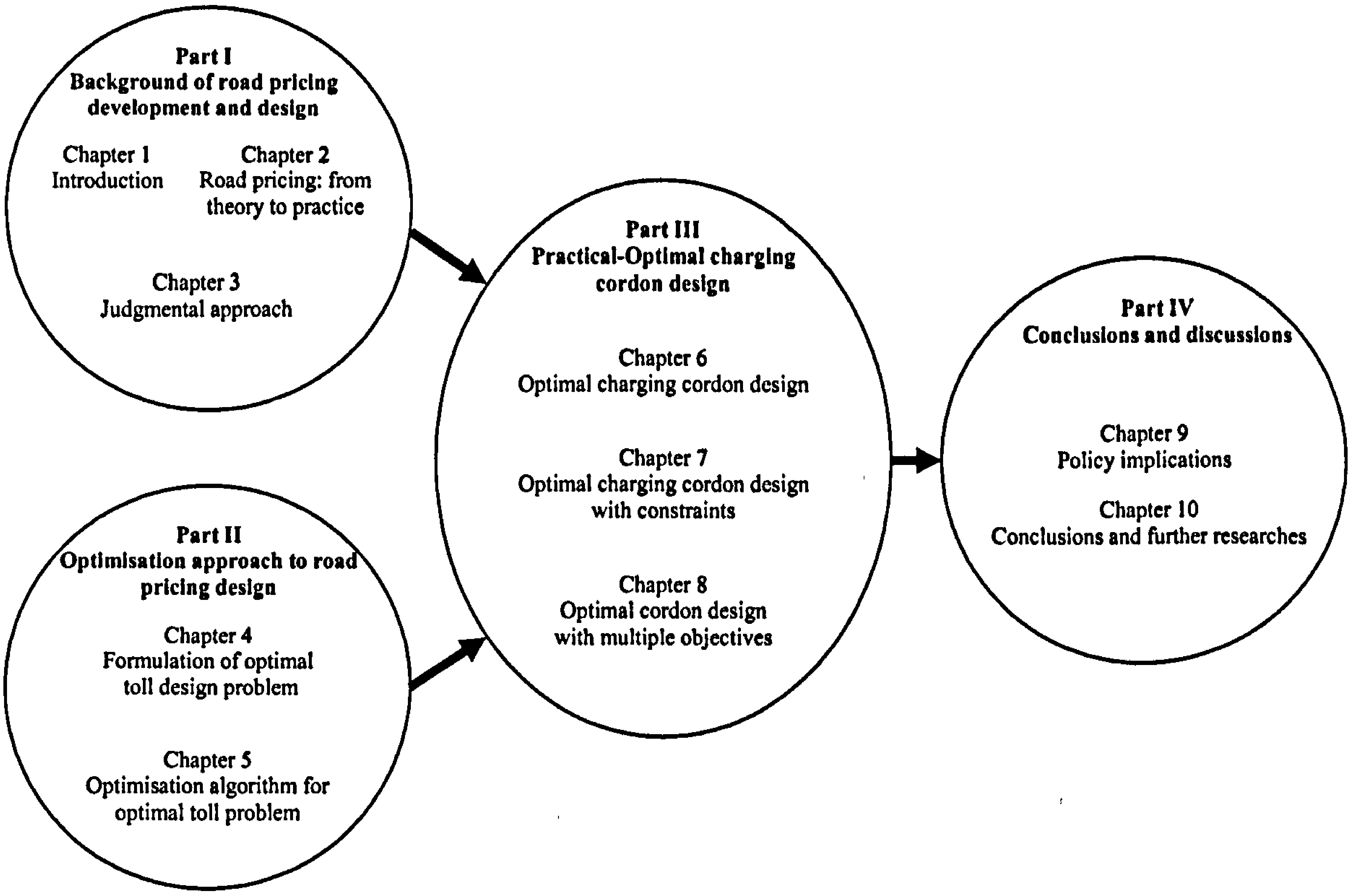


Figure 1-2 Structure of the thesis

Part II of the thesis (Chapter 4 - Chapter 5) turns the attention to the theoretical development of an optimal design for a road pricing scheme. As explained, the idea of using the model to identify an optimal transport policy is not new. The materials presented in this chapter are based on various studies in the area of network design, bilevel optimisation, and non-linear optimisation.

Chapter 4 mathematically formulates different instances of optimal road pricing design problems. The concept of user equilibrium, which represents road users' behaviour responding to the road pricing scheme, is presented and discussed in detail. The framework of a Mathematical Program with Equilibrium Constraints (MPEC) that is the core of theoretical design of an optimal road pricing scheme is also explained in this chapter. In addition, different possible performance indicators for a road pricing scheme (e.g. social welfare improvement, equity impact, net revenue) are formulated.

A brief review of various optimisation algorithms developed for solving the MPEC problem is presented in Chapter 5. A method adopted in this thesis to solve the optimal road pricing design problems is presented. The development of optimisation methods in Chapter 5 concentrates mainly on solving the pure theoretical design problems of the optimal toll scheme. Three different optimisation algorithms are developed to solve the optimal toll problem: the smoothing approach, the improved cutting plane approach, and the Genetics Algorithm based approach. However, the results from applying these design methods may not achieve the practical requirements mentioned in Chapter 3.

The main content of Part III of the thesis is the development of techniques that integrate both the practical requirements of a road pricing scheme (discussed inChapter 3) and the theoretical optimality of the design (explained in Part II).

In Chapter 6, an approach to integrating the topological constraint of a charging cordon scheme, which is identified in Chapter 3 as the most practical road pricing scheme, with the optimisation process is developed. The Genetic Algorithm based method developed can also deal with the optimal design of multi-layer charging cordons. The method is tested and evaluated with a network of Edinburgh. In particular, a comparison between the performance of the judgmental cordon and optimised cordon is made in this chapter.

The formation of a charging cordon is not only the practical design criterion arising from the surveys discussed in Chapter 3. There are other various aspects of the scheme design one must take into account during the design process. These aspects may be considered as the outcome constraints of the cordon design. Chapter 7 extends the method developed to allow the inclusion of arbitrary constraints. Four indicators are tested: net revenue, equity impact, total travel time, and total travel distance.

Based on the review and the investigation in Chapter 3, road pricing normally serves a number of objectives (e.g. increasing efficiency, reducing congestion, or generating revenue). Chapter 8 discusses different regimes of the decision making process for the multiobjective or multicriteria decision problem. An algorithm for aiding the design of a charging cordon scheme with multiple objectives is developed and tested with the Edinburgh network.

The last pat of the thesis (Part IV) is dedicated to the discussion of the optimisation results and some possible policy implications from the research conducted. Various tests are conducted in different chapters principally to demonstrate the application of the optimisation methods or frameworks. Chapter 9 will distil the results from these tests to draw some policy implications from the results and identify some common aspects of charging cordon design. Finally, Chapter 10 concludes the research and suggests some future research issues.

CHAPTER 2 ROAD PRICING: FROM THEORY TO PRACTICE

The ideas of economists and political philosophers, both when they are right and when they are wrong, are more powerful than is generally understood. Indeed, the world is ruled by little else. John Maynard Keynes

2.1 INTRODUCTION

This chapter provides an overview of the development of road pricing. Road pricing originated from the economic concept of marginal cost pricing that is aimed to rectify the failure in the transport market. Road pricing has come a long way in the transition from its original economic idea to the real world policy. The chapter opens by discussing the development of the idea of road pricing and recent research in the economic area.

Then, in Section 2.3 the focus is moved from the theoretical side of road pricing to the policy side. The objectives and different types of road pricing schemes in the real world are explained. Given the widespread instances of congestion in many cities around the world, the concept of road pricing has gained attention over the years. Many cities have considered or in some cases have implemented the road pricing scheme. The development of the road pricing scheme in these cases reveal a number of possible obstacles toward its implementation and provide useful experiences with the planning and implementation process of the scheme for other cities. Section 2.4 is therefore dedicated to reviewing the development of road pricing in different cases including London, Europe, Singapore, Hong Kong, and elsewhere.

During the transition of road pricing from theory to practice, several research questions have been raised on several issues including public acceptability, equity impact, optimal design, land use change, and technology for the implementation. These are discussed extensively in Section 2.5.

2.2 DEVELOPMENT OF THE ECONOMIC THEORY

The concept of road pricing emerged from the idea that the cost paid by the road user (called marginal private cost or perceived cost) is actually lower than the actual cost he or she imposes (called marginal social cost) (Pigou, 1920; Knight, 1924; Walters, 1961; Vickrey, 1963). As early as 1920, Pigou argued that motorists entering a crowded road network not only incurred costs to themselves but also imposed one on the other users. The cost imposed on other users is called external cost. In making their decisions, motorists consider only their own marginal private cost and will continue to take additional trips as long as the marginal private cost is less than or equal to the benefit.

Figure 2-1 shows how this situation occurs graphically. In this setting, road users are identical apart from their marginal willingness to pay for a trip, represented by the demand curve D = marginal private benefit (MPB). Due to congestion, marginal social cost (MSC) exceeds marginal private cost; the latter being equal to average social cost (ASC). The free market equilibrium outcome is t_0, and the socially optimal road use is at t_1. It is clear that at t_0, the actual costs imposed by the road users are higher than the benefit gained. By introducing the toll f^* which is equal to the marginal external congestion costs (MSC-ASC) at the optimum, the equilibrium will be driven to the socially optimal equilibrium and the welfare gain enjoyed is given by the shaded area.

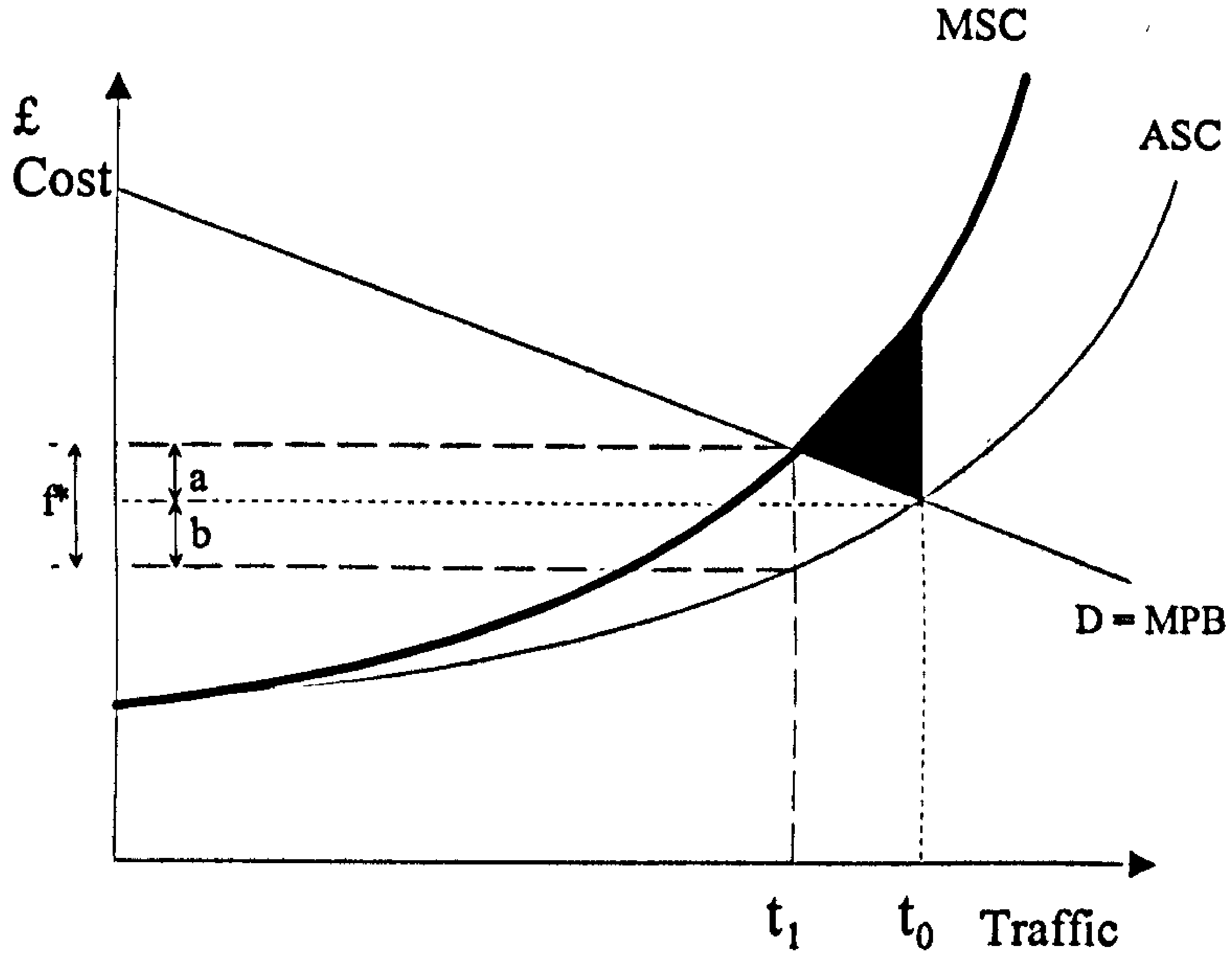


Figure 2-1 Illustration of the economic theory of marginal cost pricing

The example demonstrated in Figure 2-1 is very intuitive. However, a number of key assumptions were made to simplify the analysis. One of the key assumptions made was the absence of a network. Wardrop (1952) defined the two well-known rules of deterministic traffic assignment in a network, user optimum and system optimum. The first rule of user optimum states that under the equilibrium condition traffic arranges itself in such a way that no individual trip maker can improve his or her travel cost by switching route. The second rule of system optimum states that under the social equilibrium conditions traffic should be arranged in a congested network so that the total travel cost is minimised. In the traffic assignment literature (see, for example Sheffi, 1985), it is well known that the marginal-cost toll has been proposed to drive the user optimal toward system optimal traffic equilibrium. Namely, by levying a suitable flow-dependent congestion fee on each user using a particular link in the network, the traffic flow pattern which results from choosing cost-minimising routes between any origin-destination (O-D) pair will be a system optimum in terms of the minimisation of the total network travel cost.

Yang and Huang (1998) stated that this straightforward approach is not valid for a network with elastic demand because travel costs can be minimised by simply setting a toll so high that no travel takes place. In this case, the objective function for the optimisation program for system optimum must be to maximise the net economic benefit. They derived the mathematical program for a system optimal traffic assignment and showed that the external cost equals the derivative of the link travel cost multiplied by the link flow.

The development of the theory of marginal cost pricing relies heavily on the assumption of first-best conditions. These assumptions are not usually satisfied in the real world (Sharp, 1966). Levy-Lambert (1968) and Marchand (1968) are probably the first to study optimal congestion pricing with an untolled alternative route for a fixed time period representing the second-best condition. Their work showed how the marginal cost pricing principle must be modified to take into account an imperfect environment.

Glazer and Niskanen (1993) studied second-best optimal parking fees for a city centre where neither through-traffic nor road users with access to private parking spaces can be charged. They showed that if the second-best condition exists in road pricing, then a lump-sum parking fee can increase social welfare. On the other hand, a parking fee per unit time does not increase social welfare. The parking fee per unit time spent induces the behaviour of short stay parking and allows more persons to use parking spaces each day and hence can increase traffic. McDonald (1995) and later Verhoef (2002) used economic theory and numerical examples to investigate the question of optimal congestion pricing with an untolled route in a general network. The numerical examples showed the substantial difference between the optimal second-best tolls and the optimal first-best tolls where efficient tolls can be imposed on all routes.

Most of the studies mentioned earlier extend the concept of optimal road pricing into the network context and including the untolled alternative route to represent the second-best environment. However, the representation of the possible response of the users to the toll imposed in these studies follows a very simple assumption of the single user class and single time period.

The multiple time period and departure time choice, which are the other important characteristics of the possible response of users to the toll, has also been attracting attention from researchers in the area of second-best optimal toll design. Liu and McDonald (1999) analysed the second-best toll problem with possible departure time choices between two time periods (i.e. peak and pre-peak). The results showed that the first best charging scheme is always better than the second-best scheme in reallocating traffic volumes as well as improving social welfare. They also used a two-route example to show that the second-best optimal tolls are appreciably smaller than the first-best tolls. Chen and Bernstein (1995) considered the application of the second-best toll in the particular case of AM/PM pricing. Their research extended to only a simple case of a two-route network with an untolled route under fixed demand. The analysis of the optimal toll and subsidy took into account route choice between tolled and untolled routes. The main finding is that by carefully designing a scheme combining tolls and subsidies, a two-directional pricing scheme can be implemented with one toll point only.

Arnott *et al* (1990b) analysed user equilibrium, system optimal, and various pricing regimes for a bottleneck model of a network with two routes in parallel following Vickrey (1963). They included departure time choices into the analysis. The results showed that an optimal-time varying toll reduces the queuing without affecting the route usage. Step tolls generally yield much greater efficiency gains than uniform tolls because they reduce queuing by altering departure times.

The development of the theory of road pricing has also acknowledged the importance of the heterogeneity of the road users in which the optimal pricing applied to different groups can be defined. Small and Yan (2001) demonstrated the use of "value pricing" where there are two user groups taking in account heterogeneity in value of time. The results show that accounting for heterogeneity does improve the optimal toll design in terms of social welfare improvement. Second-best distortion in the transport market can also occur in other ways. Discussion of related second-best topics in transport can be found in Wilson (1983).

The development of the theory of optimal road pricing has not only been about the appropriate representation of the user's behaviour. The representation of the supply side is also a main research topic. Although the externalities considered in the analysis of optimal toll pricing can include several other factors, e.g. pollution and accidents, the major cause of the externalities is still congestion. Thus, to determine the optimal toll for reflecting the congestion externality, an appropriate supply model is needed to realistically represent the congestion occurring on the road as demand rises. Most of the initial analysis mentioned earlier adopted a static model that explains the congestion as a relationship with the average flows on the link. This static model of congestion simplifies the analysis of the optimal toll problem and is convenient for traditional static economic analysis.

However, the focal discussion point on the plausibility of this static speed-flow model is the phenomenon of the backward bending of the supply curve (also referred to as hypercongestion). This unwanted phenomenon of backward bending was caused mainly by the engineering definitions of flow and density used to derive the relationship of the speed-flow curve. This condition occurs when the density of the traffic becomes so high and the speed become so low (highly congested conditions) that the traffic flow (speed times density) falls below its maximum possible value. The same flow level can be produced under the normal uncongested conditions with a high speed and low density of traffic.

The discussions on the plausibility of the non-unique relationship between the speed and flow and whether using a static model to represent a dynamic traffic condition is a good approach are very critical. In addition to this engineering discussion, this backward bending phenomenon also causes a very severe question on the stability of the equilibrium between the supply and demand (due to the non-unique equilibrium point). In parallel, the question can also be asked if the economic representation of the equilibrium between demand and supply in a static manner is sufficient to study the marginal cost toll problem given the dynamic nature of the traffic and congestion.

In fact, this dynamic congestion issue is not a new topic. Vickrey (1969) analysed the optimal peak-load toll with a simple bottleneck model which can be represented in a tractable form. Several researchers since then have started to analyse the dynamics of congestion and the associated optimal peak-load toll. Arnott *et al* (1990a) include departure time choice into the analysis of dynamic pricing with a modified bottleneck model.

The key simplification of Vickrey's model is that the traffic is assumed to move freely until it meets the bottleneck and also that the traffic can move freely through the queue. This simplification was criticised to discard the shock-wave effect of the dynamics of queue development (Mun, 1999). The lack of traffic jam conditions in Vickrey's model resulting in a slow rate of queue growth (compared to the model with traffic jams) was claimed to underestimate the optimal toll. Mun (1994) introduced the traffic jam based on the theory of kinematic wave (Lighthill and Whitham, 1955) into the bottleneck model and analysed the optimal dynamic congestion toll. However, his model assumed a fixed pattern of travel demand for each time period. Later on, Mun (1999) incorporated departure time choice into the model and also analysed the effect of optimal dynamic tolls. The result showed the benefit of the introduction of a peak-load toll.

The other treatment of the dynamics of congestion is to relate the traffic jam with either the inflow at the instant the trip is started (Chu, 1995) or the outflow at the instant the trip is ended (Henderson, 1974). Obviously, neither model is able to represent the interaction of traffic on the link from different periods. Recently, there are further suggestions and developments in the area of dynamic traffic modelling in which the flow-propagation can be treated in various ways. Friesz *et al* (1993b) proposed to define the travel time of a trip as a function of the traffic volume at the instant the trip is started. Ran and Boyce (1996) adopted the travel time function with the arguments being the traffic volume on the link, inflow, and outflow at the instant of the beginning of the trip. Carey *et al* (2003) defined travel time as a function of the weighted combination of the inflow at the instant the trip is started and the outflow at the instant the trip is ended. Obviously, different models behave in a different manner and the study of the most plausible models for the study of the congestion pricing needs a comparison with real data.

Recently, Verhoef (2003a) proposed a simple car-following model for representing the flow propagation on the link as well as the development of the traffic jam and queue in continuous time. He analysed the dynamic optimal toll pricing framework with the proposed model and showed that the toll framework derived from the model proposed outperformed the toll framework based on Vickrey's model. Verhoef (2003b) later on verified the instability of the equilibrium between static demand and supply framework using the simplified car-following model. He also compared the behaviour of the simplified car-following model with the real-world traffic data.

The development of the theory of road pricing has come a long way from its original idea. It has expanded in two main branches. The first is the study of the formulation of the optimal toll problem under an imperfect environment (second-best toll problem) with different attempts to extend the properties of the demand model adopted. The second strand is the analysis of the phenomenon on the supply side which mainly involves the study of the development of congestion and queues. It can be seen that there is a rich development of the research into the theory of road pricing. In the next section, the development of the road pricing as a transport policy instrument from the practical point of view will be discussed.

2.3 ROAD PRICING AS A TRANSPORT INSTRUMENT

2.3.1 Objective of the road pricing scheme

Despite its original objective of enhancing the economic efficiency of the transport system as explained in the previous section, the practical objectives of road pricing have since been developed beyond this original focus. This is due to the acceptance of road pricing as a demand and urban management tool by transport and city planners (Jones and Hervik, 1992). The interest in road pricing has been stimulated by the desire to find a new revenue generation source for the improvement of transport infrastructure to keep up with the growth of the traffic, and by the failure of alternative policies to curb the level of traffic congestion/demand (Small and Gomez-Ibanez, 1998). Four main practical objectives for the implementation of road pricing can be defined. It is noteworthy that most of the schemes consider these objectives simultaneously with some order of preference.

(i) Reduce traffic congestion

The concept of reducing congestion is probably the first practical objective attached to the implementation of a road pricing system. This is because congestion in the traffic system is the main cause of the inefficiency of the system. Thus, the idea of using road pricing as the travel demand restraint measure was proposed. The objective of the implementation of the first Singapore ALS was purely to restrain the travel demand into the central business district in order to reduce congestion (Holland and Watson, 1978). Similarly, the feasibility study of the Hong Kong congestion charging scheme dealt mainly with the objective of congestion reduction. This is also the case for the recent congestion charging scheme in London.

However, it is noteworthy that the objective of reducing congestion may not perfectly coincide with the objective of economic efficiency. After the first implementation of the ALS in Singapore, the scheme has been regularly updated and revised to ensure the efficient usage of the road network (see the detail of the development of the Singapore road pricing scheme in Section 2.4.3).

(ii) Environmental improvement

The other main problem in most cities is the growing concern over the environmental impact of traffic congestion. Increasing worldwide concerns over environmental issues in the past decade have stimulated the discussion of the impact of transport on the local and global environmental problem. Following the Kyoto agreement on the target reduction of greenhouse gas emission, the UK Department for Transport has set a target to reduce the carbon-dioxide emission level by 20% on 1990 levels by 2010 (Department for Transport, 2000). The plan mainly aims to reduce road congestion in the inter-urban and major urban cities. Although the relationship between the level of congestion (or traffic) and the levels of different types of pollution is not clear, the interest in pricing traffic to reduce the environmental problem in the city has been spreading. For example, environmental protection was the principal focus of the scheme development in Holland and Stockholm.

(iii) Revenue generation

As mentioned earlier, the public sector has found it harder and harder to obtain sufficient funds for the improvement of transport infrastructure in order to keep up with the growth in demand. The main guaranteed benefit of road pricing is the revenue generated from the tolls. Thus, many cities and governments have been attempting to use road pricing as a financing tool for major transport investment. A good example of the application of road pricing to generate revenue is the toll ring schemes in Norway (see Section 2.4.2). The objective of revenue generation is often considered alongside the other objectives. The implementation of a road pricing system in practice imposes some implementation and operating costs on the operator. Thus, the road pricing system should be at least a self-financing scheme. Similarly, the analysis from various research suggested the importance of the revenue exploitation from the road pricing scheme to improve the transport system overall. This issue is discussed later in Section 2.5.1.

(iv) Urban planning and city protection

Although road pricing has been considered mainly as a transport policy, recently there has been a growing recognition of the benefit of road pricing as an urban management tool. Road pricing can be used to increase the accessibility of the city hence increasing the attractiveness of the city for business (although there may also be some negative impacts from the tolls imposed on the economic development) and revitalising the area. From a different perspective, the introduction of road pricing can be aimed at protecting the existing historical part of a city (e.g. the discussion on using road pricing to protect the Old Town area of the city of Edinburgh). Recently, a single toll point scheme was implemented in Durham, UK, with the clear aim to protect the peninsular area which is the main historical part of the city including a World Heritage Cathedral and Castle.

2.3.2 Possible methods for implementation

Various approaches to implementing road pricing in the real world have been proposed and studied. Three main types of charging regimes have been discussed including (i) point-based charging, (ii) area-based charging, and (iii) continuous charging.

In a point-based charging system, drivers are charged when crossing a specific point in the network. There are two types of point-based charging system, cordon-based and cellular systems (MVA, 1995). The cordon-based system involves one or more boundary lines around a specific area, and sometimes incorporates radial screen lines. Cordon based charging systems currently operate in three Norwegian cities. The scheme in Singapore has also been gradually modified to a cordon-based system. Similarly, the schemes proposed for Hong Kong and Edinburgh are also cordon-based schemes. A wide range of technologies have been adopted to implement the cordon-based system including free-flow charging system with smartcards or vehicle tags and manual tollbooths or enforcement points. Cellular systems include a system of hexagonal cells covering the charged area. The cordon based system is a very simple system but can also be modified to be a complex system by the introduction of screen lines (e.g. the proposed Hong Kong scheme). The key drawback of the point-based charging system is that it cannot impose charges on the trips inside the charged area or, more generally, between charging points.

Area-based charging scheme imposes a charge on the presence of the trip inside the charged area. The area-based charging scheme can also be referred to as a supplementary licensing system. The original ALS scheme in Singapore and the current

London congestion charging scheme are both area-based charging systems. Enforcement for the area-based scheme was operated manually in the Singapore ALS. Recently, with the improvement of technology the London congestion charging scheme has implemented a camera based enforcement system with number plate recognition technology.

The last category of charging regimes is the continuous charging system, which includes time based, distance based, and delay based charging regimes. As the names suggest, these charge the user according to the amount of time, distance, and delay on his/her journey inside the charging area. A time-based system, TIMEZONE, was proposed for Richmond, London. However, the time-based system may lead to dangerous driving behaviour in an attempt to reduce the travel time (Bonsall and Palmer, 1997). The Congestion Metering system was proposed for the city of Cambridge in the UK (Oldridge, 1990). However, the traveller may not be able to estimate the charge level for her trip prior to the end of the journey. Recently, CfIT (2002) also proposed a nation-wide distance based charging scheme in the UK.

2.4 EXPERIENCES AROUND THE WORLD[1]

2.4.1 Developments in the U.K.

The development of the idea of road pricing in the U.K started since the first major study in 1964, the Smeed Report (Ministry of Transport, 1964). The report set the tone of the discussion of the necessity of the road pricing policy for tackling the major traffic congestion problem in U.K. in that time. The detail of the design criteria from the Smeed Report will be discussed in Section 3.2 in Chapter 3. After the Smeed Report, a feasibility study of the relative benefits of road pricing and parking controls in London was commissioned, Better Use of Town Roads (Ministry of Transport, 1967). In 1974, the idea to implement road pricing in London was developed further with the proposal for Supplementary Licensing (Greater London City Council, 1974; May, 1975). The Supplementary Licensing scheme required every vehicle entering the Inner London area between 0700 and 1900 on weekdays to purchase a daily licence costing around £5 (2003 prices), with a charge of three times that level for commercial vehicles and

---

[1] Note that this section is based largely on May and Sumalee (2003)

exemptions for buses, taxis, disabled drivers and emergency vehicles. The scheme was expected to reduce the traffic entering the centre by 45% and to increase speeds by about 40% during the peak periods(May, 1975). Despite the clear benefit of the scheme predicted by the modelling study, the proposal was not pursued by the GLC mainly due to concerns over the impacts on equity and the economy. Similar proposals emerged at the same time for Bristol and York, but this was a high point in the development of road pricing, not to be regained for another twenty years.

During the early 1990s, interest in road pricing re-emerged due to the lack of success of other transport policies in tackling congestion and the need to provide a more integrated transport strategy (May and Roberts, 1995). In 1992, the London Congestion Charging Research Programme was commissioned by the UK Department for Transport (MVA, 1995). The study investigated different perspectives of the implementation of a road pricing scheme in London extensively covering the evaluation of alternative schemes, potential technologies, and administration and enforcement. The simplest scheme under study was a single cordon charge around Central London whereas the most complex scheme investigated involved three cordons and screen lines. Again, the study confirmed the benefit of introducing road pricing in London. It should also be noted that in that time there did not exist any legislation for the introduction of road pricing in the U.K. cities. During the same period, several cities in the U.K. including Edinburgh (May *et al*, 1992), Bristol, and Leicester also investigated the feasibility and benefit of the implementation of a road pricing scheme in their cities as part of an integrated strategy.

In 1998, based upon the evidence of the studies in the 1990s the incoming Labour government decided to provide local authorities with the power to implement congestion charging schemes and workplace parking levies, and to retain and use the revenues collected from these schemes for other transport projects for the next ten years after the scheme implementation (DETR, 1998). In order to stimulate and support the development of congestion charging schemes, the government set up the Congestion Charging Development Partnership. Initially, almost 30 local authorities who were interested in implementing road pricing or workplace parking levies joined the partnership.

In London, the development of road pricing was stimulated by the change of the local political institution. Legislation in 1999 established the Greater London Authority. After the establishment of the GLA, the first Mayor was elected in 2000. Prior to the Mayoral election, a major research programme Road Charging Options for London, was conducted by an independent group of transport professionals (ROCOL working group) to inform the newly elected Mayor. The Mayor produced the Transport Strategy for London which included congestion charging as the main policy to tackle the transport problem in London based on the ROCOL study (GLA, 2001).

In February 2003, the congestion charging scheme commenced in London. The scheme is an area licensing scheme covering Central London. The £5 charge is applied to trips travelling inside the charging zone between 7am to 6.30pm weekdays (taxis and motorcycles are exempted). Residents inside the charging zone receive a 90% discount. Transport for London (TfL) estimated, prior to the implementation of the scheme, a reduction in traffic entering the charging zone of around 17-28%. The recent impact monitoring report suggested a 18% reduction of the trips entering the zone (TfL, 2004).

During the feasibility study of the scheme, the scheme was expected to generate net revenues of £120 million in 2003/04 and £130 million in subsequent years (the operating costs per year are about £97 million). However, the latest estimates are that the net revenue will be around £68 million in 2003/04 and £80 million to £100 million in subsequent years (TfL, 2003). The main reasons for the lower level of the net revenues generated are an over-estimate of the number of vehicles entering the charging zone from the modelling forecast, the number of exempt and discounted vehicles being higher than expected, fewer commercial vehicles than expected, and higher levels of evasion of the system. Currently, there has been a discussion on the extension of the scheme to the west part of Inner London. The main reason for this plan is the congestion problem remaining in that area.

The other active developments of road pricing in the U.K. includes the recent implementation of the single toll point charging scheme in Durham, the discussion of the double-toll ring scheme in Edinburgh and the development of a single cordon proposal in Bristol. Despite the success of the London congestion charging scheme, the climate for development of the road pricing scheme in the U.K. has not been as active as expected. This is because London is not typical of cities in the U.K. and hence the success of the scheme in London may not be transferable to other cities. The scheme in Edinburgh should have a major impact on the development of road pricing in the U.K. It covers a wider area of the city and the city itself is in a similar environment to other major cities in U.K compared to London. The current debate over the scheme in Edinburgh is the exemption of residents of the city of Edinburgh from the outer cordon charge. This is considered as a major equity problem by the residents outside the city.

While most of the interest in the UK has inevitably focused on urban congestion charging, recent reports have advocated the use of distance-based charges nationally on congested roads, offset by the abolition of the annual vehicle tax and some reduction in fuel taxes (CfIT, 2002). A system of this kind is scheduled to come into operation for commercial vehicles in 2006, and a government field trial of the technology, based in Leeds, is expected to start shortly after some considerable delay.

2.4.2 Developments in other countries in Europe

Norway

In Norway, road pricing has long been used as a supplementary fiscal instrument to raise finance for road projects. Currently, 25% of the total annual budget for road construction in Norway comes from the road pricing schemes around the country (Odeck and Brathen, 2002). Most of the road pricing schemes impose tolls on particular sections of trunk roads, tunnels, or bridges. Only five of them are urban charging cordon schemes (or toll rings): in Bergen, Oslo, Trondheim, Stavenger, and Kristiansand. However, recently discussions have taken place concerning the modification of the current toll financing schemes to congestion charging schemes in Bergen, Oslo, and Trondheim.

The Bergen toll ring was introduced in 1986 aiming to directly raise finance for completing the planned road system. In 1990 the capital city of Norway, Oslo, also introduced an urban toll ring, to finance a new tunnel under the city centre. The implementation of the tolls in both cases was timed to coincide with the opening of the new tunnel and bypass projects financed by toll revenues. In 1992, a toll ring was implemented in Trondheim, which has been gradually developed over the years since its introduction. An *'amputated'* toll ring with only two toll plazas was in operation from 1992-1996 in Kristiansand. A new package and toll charge period were recently agreed to fund the construction of the new trunk road (E18) and two tunnels through Kristiansand. In 2001 Stavanger implemented a city toll "ring". The toll will be in operation for 10 years to finance the new road and other transport projects. Table 2-1 summarises the characteristics of the schemes in these five cases.

City	Bergen	Oslo	Trondheim	Kristiansand	Stavanger
population	213,000	456,000	138,000	70,000	103,000
Starting Date	Jan, 1986	Feb, 1990	Oct, 1991	April, 1992	April 2001
Number of toll stations	7	19	22	5	21
Charging regime	Uniform charge	Uniform charge	Peak and Off peak charge	Uniform charge	Peak and Off peak charge[2]
Entry charge for small vehicle[3] (NOK)	10	15	15 (for all period for manual payment[4])	10	10 for Peak 5 for Off peak
Charging period	Weekday 6am-10pm	All days All hours	Weekday 6am-6pm	Weekday 6am-6pm	Weekday 6am-6pm
Discount	Discount for monthly subscriptions	Discount for prepaid tickets	Discounts for users of electronic systems.	Discount for monthly subscriptions	Several advance payment discounts with AutoPass
Annual gross revenues (NOK millions)	156	1,046	168	95	80
Annual operating costs (NOK millions)	30	103	17	20	21

Table 2-1 Key characteristics of the Norwegian toll rings (Source: May and Sumalee, 2003)

Given the original objective of raising revenues, the lower toll level in all schemes only reduced the traffic slightly (around 6-7% for Bergen, 3-4% in Oslo, and 10% in Trondheim during the charged periods). Originally, in Bergen the toll revenues collected were only used for road projects. A new agreement was reached in 2002 for

---

[2] Peak period: 7am-9am and 2pm-5pm; off peak period: other period between 6am-6pm.

[3] Heavy vehicles are charged double price

[4] For prepayment of 6000 NOK, 9 NOK between 6am-10am and 6 NOK between 10am-6pm; for prepayment of 3000 NOK, 10.5 NOK between 6am-10am and 7.5 NOK between 10am-6pm; for prepayment of 1000 NOK, 12 NOK between 6am-10am and 9 NOK between 10am-6pm.

maintaining the toll ring system until 2011 with the basic toll levels increased to 15 NOK from 2004 onwards (which coincides with the implementation of electronic collection), only 45% of the revenues will be allocated to road investment, and the scheme will be refocused as a congestion charging system (Ramjerdi *et al*, 2004). In Oslo, Trondheim, Kristiansand, and Stavanger, the revenues will help finance road projects, public transport improvement, and other safety instruments. New toll ring schemes are also underway. In 2003, the Namdal project (in the city of Namso) started which is claimed to be the smallest toll ring in the world (only two toll points). Tønsberg will make a decision on the introduction of a toll ring by 2004.

The toll ring system in Norway is currently at a crossroads. Most of the projects around the country were originally initiated to finance major local transport schemes (mostly road transport infrastructure). The agreements for many existing schemes are near to the end or already terminated (in the case of Bergen). A decision on the future of the toll rings has to be made. At the national level in Norway a new law on tolling and road pricing has just been sanctioned by the Parliament. Through this law road user charging is accepted as a means both for revenue raising and for demand management, but the two objectives can never be mixed. This means that today's tolling systems must be terminated before any urban pricing scheme can be considered. Public acceptance of these changes is also uncertain. While 54% opposed Bergen's toll ring before its implementation, that had fallen to 37% a year later (Odeck and Brathen, 2002). It is not clear whether toll rings designed for congestion charging will attract such majority support.

Sweden

Sweden has had an interest in restraining traffic, particularly in Stockholm and Gothenburg, since the 1980s. Its main focus has been protection of the environment, although relief of congestion has also been an issue. The most significant proposal for Stockholm emerged in 1991 as part of the Dennis agreement (Gomez-Ibanez and Small, 1994). The Dennis package involved relieving traffic problems in the inner city by improving public transport, building an inner ring road and a tolled western bypass, and introducing a toll ring, just outside the inner ring road. Tolls would have been around $2 at current prices, with the possibility of variations by time of day and by standard of emission controls. With the outer bypass tolls, they would have been designed to provide the main source of finance for the investments. While the proposals initially had the support of all the main political parties, it soon became clear that both the inner ring road and the toll ring were highly controversial, and the proposals were dropped in 1997. However, other agencies, including the Swedish Society for Nature Conservation, the Swedish Institute for Transport and Communications Analysis and the Swedish Environmental Protection Agency, have since submitted proposals for somewhat similar pricing schemes. Most recently, the Swedish National Road Administration has published a review of the options for road pricing in urban areas (Eliasson and Lundberg, 2003). While this does not make specific proposals, it is one of the most comprehensive summaries of successes and failures in road pricing currently available.

The Netherlands

During the late 1980s, the Dutch Government proposed the introduction of a large multiple cordon-based road pricing system called *rekening rijden* ('road pricing') for the Randstad region (including Amsterdam, Rotterdam, The Hague, and Utrecht, plus part of the province of Noord-Brabant). The main objective of this proposed scheme was to manage travel demand, and hence to reduce congestion. Other objectives of the scheme were to decrease environmental pollution and generate funds to finance new infrastructure. Due to public opposition the proposal was not pursued. In 1991, a more conventional form of road toll using toll plazas (*tollpleinen*) was proposed. The objective of the scheme was redefined to solely raise money for road infrastructure. However, due to the potential disruption of the traffic caused by the stop-and-go operation of the toll plazas and the amount of land required for implementation, the proposal was rejected.

In 1992, a reduced scope proposal involving a system of supplementary licensing for motorists using the main road network during peak periods (*spitsvignet*) was discussed. The rush hour motorists would have been charged a fixed amount toll to travel during peak hours regardless of the area. The charge would be about $2.85 per day (1992 prices) applied during the morning rush hour 6am-10am. However, the proposal was not approved after a new government was elected in 1994. Boot *et al* (1999) suggested that the most important reason for the failure of these earlier proposals was political acceptability.

Subsequently, in October 1994 the Dutch parliament agreed in principle and strongly proposed the implementation of a revised form of *rekening rijden* (referred to as 'congestion charging') which would be a system of electronic toll cordons around the four main cities in the Randstad area starting in year 2001 (Dutch Minister of Transport, 1995; Transport, 1995). The charge would be in operation during the morning rush hour (7am-9am) on weekdays. The objectives of this proposal were to improve accessibility of the economic centres.

In 2001, congestion charging became a major political debate in the Netherlands. The proposal of rekening rijden was opposed by several interest groups. The main objection was that the authorities failed to provide an alternative for those who were obliged to travel by car during the proposed charging period. The government is now considering an alternative proposal for a *Mobimeter* ('kilometre charging') system. The idea was supported by the successful development of the technology for the kilometre charging system. In addition, the policy could well fit in with the European Commission White Paper which proposed a kilometre charging system as a good instrument for transport pricing in Europe. The system is expected to be fully operational by 2006. The system will be a non-differentiating kilometre charge first but the possibilities of differentiating the charge in relation to congestion will be discussed further.

The barriers to the success of the implementation of congestion charging in the Netherlands have been politics and technology. The success of the recent proposal for a kilometre charge will rely heavily on the reliability and capability of charging technology. However, the greatest barrier for further progress still seems to be a political one, and is closely linked to the issue of public acceptability.

Germany

The key development of road user charging in Germany is the implementation of interurban freight charging. Since April 2001 there has been a standard emission-related tariff for motorway tolls applicable to heavy goods vehicles, jointly implemented by

Belgium, Denmark, Germany, Luxembourg, the Netherlands and Sweden (Eurovignette). The current system of Eurovignette charges all heavy goods vehicles with a weight more than 12 tonnes (except buses, coaches, and specialist vehicles) for using the road network in any of these countries. The charges are varied according to the number of axles and engine emission standards.

Germany is facing the problem of continuing freight traffic growth as the consequence of the Single European Market and the enlargement of the EU to the east and globalisation. Freight traffic is forecast to grow by 64% before 2015. The German government aims to tackle the problem by creating an efficient transport infrastructure to accommodate the growth in traffic demand, improve the rail freight network, and create fair competition between different modes. One key strategy is to rectify the price ratio between the rail and road sectors. Thus, after a long discussion the introduction of distance-related charges for the use of the motorway system by heavy goods vehicles (HGVs) was approved by the government in April 2002. The act allows the introduction of distance based charging on the motorway network and some parts of the federal highways (for safety reasons mainly) and the toll revenues can be used for infrastructure projects.

The toll system will be changed probably in late 2004 from the old Eurovignette system to the kilometre charge system. The charge will still be differentiated according to engine emission standard and number of axles. It will replace the Eurovignette and some part of the fuel duty. The charge is expected to vary between 0.10 Euro/km and 0.17 Euro/km, and is in line with the EU directive 1999/62/EC (Commission of European Community, 1999). Drivers will have two alternative charging options. The first option is the automatic charging option which is for vehicles equipped with an On-Board Unit (OBU). This automatic electronic system can be located exactly by satellite and continually transmits the position of the vehicle, the company and vehicle data, as well as the kilometres travelled on charged roads to a central computer. An automatic procedure then charges the toll in arrears to a pre-selected payment partner. The second option is the manual pre-booking system. The manual procedure requires that the driver or the vehicle's owner stipulates a route in advance and 'buys' the route at one of the toll terminals or via the Internet before the journey.

2.4.3 Developments in Singapore

Given the limited land space, the Singapore government has foreseen the possible severe impact of traffic congestion on the development of the country (Foo, 2000). The government has been trying to control the level of car traffic in the network through various generations and combinations of pricing measures over the last 30 years. Two means of controlling car travel demand have been adopted: the control of vehicle ownership and the restraint of vehicle usage.

A tax on new vehicle registrations was introduced in 1972, and tax rates were subsequently increased as a means of controlling ownership. However, there was concern that the tax was inflexible, and that it was not imposing sufficient control. In 1990, the government introduced a unique form of vehicle ownership control, the Vehicle Quota System (VQS), in which a quota for new vehicles in any month is determined to match an approved overall growth rate of 3% p.a. and the payment is determined by a bidding system. After the implementation of the VQS, the average annual motor growth rate was decreased to around 2.83 per cent from 4.4 per cent. The VQS also generated a substantial amount of revenue for the government (around $1.8 billion in 1994 alone).

Although additional taxes had been in place since 1972, the Singapore government was not satisfied with the effectiveness of this measure in curbing congestion. In 1975, Singapore introduced the world's first urban road pricing scheme, the Area Licensing Scheme (ALS), to increase the incentive for car users to switch to public transport. The original ALS was a single cordon covering the central business district (CBD) of Singapore, called the Restricted Zone (RZ). Under the ALS, a permit had to be purchased to travel into the RZ by car during peak traffic periods, with exemptions for those with four or more people (Holland and Watson, 1978). Enforcement was based on manual operation by police personnel located at each of the entry points. The morning peak car traffic volume entering the RZ in 1992 was approximately a half of the level 17 years before the ALS was introduced. Speeds had increased by 20% and accidents had fallen by 25% (Menon, 2000). Public transport's share for working trips increased from 33% in 1974 to 67% in 1992.

Initially, the charge structure was a flat rate charge of S$3 for travelling inside the RZ in the AM peak period (7:30-9:30 a.m.) on Monday to Saturday. However, three weeks later the charging hours were extended until 10:15am in response to the substantial increase in traffic volume entering the RZ just after 9:30am (Chin, 2002). The charge was then increased to S$4 and S$5 in 1976 and 1980 respectively. Gradually, the structure of the charge and charging period was modified to increase the effectiveness of the scheme. In 1989, the charge period was extended into the PM peak (4:30pm – 7:00pm) with a charge level of S$3. The charge period was extended again to the whole day from Monday to Friday in 1994 with the same charge level of S$3. The ALS was considered successful and it was also claimed that there was no significant impact on businesses inside the RZ (Seik, 1998).

Nevertheless, the original ALS also had unintended adverse effects such as congestion on feeder roads and expressways leading to the CBD (Goh, 2002). The government decided to introduce the Road Pricing Scheme (RPS) to regulate traffic on the expressways and feeder roads in 1995. The RPS (manually enforced) was implemented on the three main expressways heading to the CBD with congestion tolls to pass defined points. About 16% of motorists stopped using the expressways during the RPS operation hours (between 7:30am and 9:30am). However, the ALS and RP schemes were claimed to cause under-utilisation of the roads within the CBD and not to be able to deter the congestion outside the RZ and RPS. In addition, the manual operation of both systems was too labour-intensive and not flexible enough to permit the future modification of the scheme.

In 1998, the electronic road pricing system (ERP) was implemented. The ERP cordon covered a similar area to the original RZ of the ALS. However, the charging is imposed on a per crossing basis which is different from the original operation of ALS. An incomplete second cordon has since been implemented. The ERP charge rates are set based on the types of vehicle (including motorcycles). The charges are also differentiated according to location of crossing, day, and time of day. The road authority in Singapore reviews speeds on the expressways and roads, where the ERP system is in operation, quarterly. After the review, the ERP rates are then adjusted to maintain average traffic speed on expressways and roads inside RZ at 45-65 km/hr and 20-30 km/hr respectively.

Immediately after the implementation of the ERP, the traffic volume on the heavily congested roads fell by 17% from the condition during the operation of ALS. Traffic volume into the CBD decreased by 10-15% compared to the condition during the ALS operation (Chin 2002). ERP has been effective in maintaining a speed range of 45 to 65 km/hr for expressways and 20 to 30 km/hr for major roads as intended. The estimated monthly revenue from the system is S$3.4 million which is substantially lower than the revenue collected from the old ALS and RPS schemes, which was about S$5.8 million/month (Goh, 2002). The change of the fundamental principle of charging from ALS which allowed multiple entries for the whole day to the ERP which charges per crossing is the reason for the significant drop in the demand despite the lower charge rates.

The Singapore government has adopted a "stick and carrot" policy where a substantial amount of money has been invested in improving the public transport system. After gaining sufficient revenues from ALS, in 1988 the Singapore government decided to develop the heavy rail Mass Rapid Transit (MRT) and later on a light rail network (initiated in 1999). The development of public transport has enhanced the increase in modal split of public transport which increased from 46% in 1976 to 70% of in 1991 of all journey-to-work trips to the CBD.

The Singapore LTA plans to modify the charging area and charge levels to achieve better utilisation of the road network whilst maintaining an acceptable level of service. Although there has been a wide range of well documented papers on the success and implementation path of road pricing in Singapore, there has been very little discussion on public responses. The stable political climate in Singapore has supported the government and LTA in adopting a very aggressive transport policy over the last three decades. Despite all the successes, questions have been asked on the extent of decentralisation of the city and economic impact of the cost of the journey to work (Phang, 1993; Willoughby, 2001).

2.4.4 Development of ERP in Hong Kong

In 1982, the Hong Kong government decided to adopt fiscal controls to contain traffic. Particular measures introduced were trebling the annual fee for private cars and doubling the fuel tax and the registration fee for new cars. As a result of the vehicle ownership restraint, private vehicle ownership decreased from 211,000 in 1981 to 170,000 in 1984. However, the level of congestion was only reduced in the least congested (low income) areas and in the same time rose in the most congested areas (Dawson and Brown, 1985). Private car and taxi use still remained high, particularly during peak periods (Lewis, 1993).

In response to this failure, in 1983 the Hong Kong government decided to commission a two-year investigation of the viability of introducing a road user charging scheme using an electronic road pricing system (ERP). Hong Kong chose not to adopt a low-tech option like the ALS in Singapore on the basis that it would be too liable to fraud and require a considerable amount of enforcement (Borins, 1988). The principles of the proposed ERP scheme were similar to that of the current ERP system in Singapore (with a charge per crossing). Three schemes were designed with different locations of charging cordons, screen lines and charge structures. The schemes were primarily designed to cover the most congested areas, Hong Kong Island and Kowloon. The charge structure was planned to vary by time period and area. The combinations of different charging cordons and screen lines with different charging structures were designed following the idea of a theoretical optimum (Dawson and Catling, 1986).

The system proposed in the 1983 study was based on automatic vehicle identification (AVI) with a passive electronic number plate (ENO) mounted underneath the vehicle. At the charging points, inductive power and receiver loops installed underneath the road pavement surface would be used to detect and identify the vehicle crossing the point. The information on crossing vehicles and their crossing times would then be transmitted from the roadside computer to the main accounting and billing system. The motorists crossing the charging points would then receive a bill on a monthly basis. Enforcement would be conducted via closed-circuit television which would record the rear number plates of the vehicles. The technological tests with around 2,600 cars confirmed a very high reliability rate for the system. The proposed ERP system was expected to decrease the traffic volume by at least 20% during the peak hours and the capital cost of the scheme was estimated to be around $30 million (in 1983) (Borins, 1988).

Following the success of the technological trial and potential positive outcome of the ERP system the Hong Kong government decided to consult the district boards, which represented the public. The government faced two main arguments: the need for road pricing given the scale of the congestion problem and the potential for invasion of privacy. In early June 1985, the proposal of the ERP scheme was unanimously turned down by the district boards (Leung and Liu, 1985).

In 1994, the Hong Kong government revived the idea of tackling traffic congestion by road pricing. The government commissioned a major feasibility study, which began in March 1997, with the objective of examining the practicality of implementing ERP in Hong Kong. Various technological alternatives were considered including the Dedicated Short Range Communication (DSRC) system as currently operated in Singapore and the Vehicle Positioning System (VPS) based on Global Positioning System (GPS). A cordon-based charging scheme was still the preferred alternative for the charging regime.

Similarly to the scheme design in 1983, the charging zone would cover the most congested areas of Hong Kong and be operated on a directional and time period basis. The initial suggestion was that the peak period charge would be from 8:00am to 9:00am and from 5:30pm and 7:00pm. A slightly lower charge would be applied during the inter-peak hours. The charge rate would be set to maintain a target speed of 20km/hr. It was estimated that the implementation of this proposed ERP scheme would reduce car trips entering the charging zones by up to 50%, with 40% diverting to public transport and 10% changing travel time. In order to rectify the failure of the first proposal, there was a well-planned public consultation programme to allow public input into the development of the scheme.

Technology trials were conducted in late 1998 with both DSRC and VPS technologies. The results showed that both DSRC and VPS could be adopted in Hong Kong and the privacy issue could be overcome. However, in 2001 the government concluded that based on the feasibility study report in 1999 there were no transport and environmental grounds to justify electronic road pricing(Legislative Council, 2001). Therefore, the government decided not to pursue the implementation of the ERP scheme, despite the promising results of the technological trials. Although the technological barrier in relation to the privacy issue has been overcome, the question of the political and public acceptability of ERP still remains.

2.4.5 Developments elsewhere in the world

There are a few road pricing proposals elsewhere in the world, and most of these are using road pricing as an infrastructure financing tool rather than as a congestion charging measure. In Australia, several high-technology tolling systems are in place: a series of tolled motorways, bridges, and tunnels in Sydney, City Link in Melbourne, Gateway Bridge/motorway and Logan motorway in Brisbane. The interesting issue for Australia is the national policy to allow a customer of one toll road operator to be able to *"seamlessly"* use other toll road systems. In the recent AusLink Green Paper, the possibility of moving the existing toll financing scheme to a congestion charging scheme is mentioned (DOTARS, 2002). A road user charging system for heavy goods vehicles based on variable weight and distance (a mass-distance regime) was also referred to as an alternative.

In New Zealand, the paper based Road User Charges for heavy goods vehicles, introduced in 1977, is a weight-distance tax which relies on vehicle distance measurement devices. The purpose of this system is to recover road costs from heavy vehicles. In 2002, the government announced its intent to introduce an electronic road user charging system (eRUC) in order to increase fairness and efficiency of the charging system to vehicle operators. Migration from the paper based system to the new electronic system will be voluntary. Currently, a feasibility study is being carried out to investigate the business case and functionality design.

There have also been various road pricing proposals in South America. There was an early feasibility study of implementing road pricing in Caracas, Venezuela. More recently, the city of Santiago in Chile has outlined a plan to implement an urban road pricing scheme.

In Asia, the rapid growth of economy in this region has catalysed the growth of traffic and vehicle ownership. In Seoul, after several decades of rapid growth in car usage, the Seoul Metropolitan Government (SMG) has taken several measures to reduce congestion in the inner city and increase the mode share of public transport. Since 1993, the government started investigating different traffic demand management techniques through various fiscal tools including congestion charging.

In 1996, the SMG implemented congestion tolls (around $2.20 for both directions) on two main tunnels which link the downtown area to the Southern part of the city (Hwang *et al*, 1999). The objectives of this implementation are three fold: reducing low occupancy vehicles, raising revenues for transport related projects, and assessing the effectiveness of the pricing technique. Private vehicle cars with three or more passengers are exempted from the tolls. Traffic volume decreased by 20% in the first two years after the operation. Average traffic speed increased by 10 km/hr. A proposal for expanding the current congestion charging system in Seoul has been developed based on point charging. However, this expansion of congestion charging has not been implemented to date due to political concerns.

Following the success of the ALS implementation in Singapore, in the 1970s the World Bank funded feasibility studies of implementing a similar scheme in Kuala Lumpur (Malaysia) and Bangkok (Thailand). Although the studies strongly supported the implementation of the schemes in both cities, initial set backs have delayed implementation. In Kuala Lumpur, gantries were already installed in various points around the charging zone boundary. However, the operation of the ALS was ultimately deferred by the government.

The reasons given were that the city needed to improve public transport and complete the inner ring road as an alternative for through traffic first (Armstrong-Wright, 1986). In addition, it was claimed that the success of other road improvements in that time was able to sufficiently reduce the congestion problem in the central area. Interestingly, it was the same political decision makers who both approved the initial plan and deferred it later on.

In Bangkok, the proposal for the implementation of ALS was immediately rejected by the government due to political concerns. There has been no implementation of any form of congestion charging systems in these cities to date. On the other hand, Thailand, Malaysia, and also other countries in this region (including Philippines, China, and Taiwan) have been progressive in using road pricing as a means to finance road infrastructure project. There are various road toll projects both in urban and inter-urban contexts in these countries with the sole objective of financing the road construction.

In Japan, the Tokyo Metropolitan Government (TMG) developed the Transport Demand Management (TDM) Tokyo Action Plan in 2000. The TDM Tokyo Action Plan envisages future implementation of road pricing in the centre of the city. The TMG set up a committee to look at the possible implementation of a road pricing scheme. In 2001, the committee produced a report that proposes four different charging cordon designs. In early 2001, an electronic toll collection system was introduced in the Tokyo area; it was expanded to cover over 600 existing toll points and went nationwide in November 2001. The initial purpose of this electronic toll system was for financing, but the emerging policy in Japan is to differentially price the roads to reflect congestion and environmental impacts. Currently, experiments for congestion and environmental charging are being conducted in various locations.

2.5 RESEARCH ISSUES[5]

2.5.1 Public and political acceptability

Road pricing aims to charge for the use of roads which the users could use for free at present. Thus, problems with public acceptability of the idea are inevitable. As a result, a major obstacle to the implementation of the road pricing scheme is the political acceptability induced by the fear of public opposition.

Academics have recognised this problem and a number of surveys have been conducted to gauge the attitude of the public toward the implementation of road pricing. A total of 29 surveys were conducted in the U.K. between 1989 and 2002, and the results from these surveys show a wide range of the level of acceptability of the public to the

---

[5] Note that this section is based largely on May and Sumalee (2003)

implementation of road pricing (ranging from 8% to 76%) (Jaensirisak *et al*, 2004). This implies that the way in which people decide to accept or oppose the idea of road pricing must be influenced by other additional factors. Based on the research, five factors have been identified as the main causes of the acceptability of an individual toward the idea of road pricing.

The first factor is the perception of the seriousness of the transport problem in the city and the perceived effectiveness of the road pricing in tackling those problems (PATS Consortium, 2001). People with concern over the quality of the environment and negative effects of traffic are more likely to accept charging than others. Bartley (1995) reported the survey result as a part of the MIRO project on the relationship between the perception of the transport problem and the acceptability of congestion charging. The main finding is that the acceptability is loosely related to the perception of the problem. However, there exists some inconsistency amongst the empirical results on the inter-relationship between the acceptability and the perception of transport problems. While several studies found a strong relationship between the acceptability and perception of problem, other results suggest an important distinction between the effect of the perception of the environmental problem and the perception of the congestion problem on the acceptability of road pricing (Harsman *et al*, 2000). Harsman *et al* (2000), based on their survey, reported that people are more willing to accept road pricing as a means to tackle environmental problems.

The second factor is the hypothecation of the revenue collected from road pricing. Jones (1998) concluded that, *"Most professional and governmental bodies in the UK now accept that hypothecation of revenues will be part of the price that will have to be paid to gain sufficient public support for urban road pricing to ensure its introduction in this country"*. There seems to be a consensus amongst academics, practitioners, and politicians that revenue hypothecation is the key ingredient to the success of road pricing. However, the detail on the ways in which the revenue should be spent is still inconsistent. The revenue can be used to improve the road network, improve other modes (rail or bus), or reduce general motoring taxation. Rietveld and Verhoef (1998), based on a questionnaire survey in the Netherlands, reported the popularity of road investment as a preferred option for revenue usage.

The third factor is the perception of the user on the issue of freedom and fairness (Jones, 1998; PATS Consortium, 2001) and concerns over equity issues (Giuliano, 1992; Langmyhr, 1997). These issues are discussed more extensively in the next section. The fourth factor is related to the personal characteristics and constraints on the transport choices. These may include variation in gender, age, education, income level, access to car/public transport, household type, life style, number of children, and location of household. Nevertheless, Schade and Schlag (2000) reported the low influence of these characteristics over the acceptability of the road pricing scheme compared to other attitudinal factors.

The last factor is the features of the charging system. Sheldon *et al* (1993) stated that a more complex cordon system would experience difficulties in implementation due to the opposition of the public. A simple system tends to be more acceptable. Politically, Rom (1994) also suggested that a congestion charging programme which did not rely on complex strategies of implementation would be more politically attractive. Similarly, Bonsall and Cho (1999), based on their survey, reported the preference of users for a simple charging system. However, Schlag and Schade (2000) found similarity of acceptance between distance based, congestion based and cordon based pricing schemes. In relation to the features of the charging scheme, the toll level obviously influences the level of public acceptability. Both charging structure and toll level define the level of impact on the population. Jaensirisak *et al* (2003), based on a questionnaire survey, established the relationship between scheme design, toll level and the acceptability of a road pricing scheme.

Clearly, the issue of public acceptability will still be a major research topic for the development of a successful road pricing scheme. Many questions are still waiting for answers, e.g. the way the revenue should be spent to maximise the public acceptability, the detailed relationship between scheme design and acceptability, and the clear relationship between the perception of transport problem and acceptability. Nevertheless, significant progress has been made and has contributed to the successful implementation and policy development of many charging schemes around the world.

2.5.2 Equity impact and social exclusion

Equity issues have been a focus of concern for some considerable time (Small, 1983; Else, 1986; Cohen, 1987). Various definitions and dimensions of equity as the result of road pricing have been suggested. Viegas (2001) and Jones (2002) pointed out that the definition of equity in transport is largely concerning fairness of the right of access to transport infrastructure for different groups of people. This raises the question of whether road pricing is a fair allocation mechanism amongst different groups of individuals. Giuliano (1994) suggested that the equity issue in road pricing must consider both the distribution of benefits associated with reduced congestion (including side-benefits such as pollution reduction and improved public transport service) and the distribution of costs needed to achieve the congestion benefits. Schade and Schlag (2003) suggested the psychological view on the issue with reference to the term of 'justice' which may be different from the idea of a fair allocation mechanism.

Regardless of the exact definition of equity, for analytical purposes it is necessary to define groups of potential winners and losers from road pricing (Langmyhr, 1997). Mainly, there are two dimensions of equity: vertical and horizontal. The vertical dimension of the equity issue concerns the unequal impact from the scheme across different groups of the population segregated by income and socio-economic characteristics. For instance, one may argue that the implementation of a road user charging systems will benefit the rich and exclude the poor (or lower income level group). The vertical equity analysis is mostly associated with the protection of those in worst conditions(PATS Consortium, 2001). Jones (2002) referred to the vertical equity as social equity. The horizontal dimension of the equity impact is referred to as the spatial equity impact or territorial equity. The horizontal equity impact can be described as the impact on the population living in different parts of a certain area. If the scheme benefits only a small group of people from some areas, but the rest of the population experience a decline in the social welfare, the scheme can be argued as inequitable.

Early attempts at dealing with the equity issue mainly involved analysing the impact of road pricing on vertical equity (See Gomez-Ibanez, 1992; Giuliano, 1994; Anderson and Mohring, 1995; Langmyhr, 1997; Fridstrom *et al*, 2000). A general conclusion from various research studies was that low-income car users or less-flexible car users (e.g.

based on gender or flexibility of working schedule) are likely to be the worst-off groups as a result of road pricing. If revenues are not redistributed in any way, road pricing generally result in gains for higher-income groups and losses for lower-income groups(Else, 1986; Cohen, 1987). The way the revenues are distributed has a significant impact on the equity issue (Small, 1992; Giuliano, 1994; Fridstrom *et al*, 2000).

Some research has looked at the horizontal dimension of equity. Fridstrøm *et al* (2000) analysed the spatial impact of road pricing cordons using spatial accessibility for each zone segregated by modes as the indicator. They suggested that the main adverse impact of a charging cordon is its boundary effect which also depends on the actual design of the scheme. A small cordon would affect residents inside the cordon most whereas those outside the cordon are the main victims for a wider cordon scheme. In the study of the Singapore ALS, Holland and Watson (1978) indicated that the cordon gave more advantage to the commercial firms outside the cordon. Obviously, this problem may be eliminated by the introduction of a different charging regime such as time-based, distance-based, or delay-based regimes (Jones, 2002). Halden (2003) also used the accessibility ratio between car and non-car from different zones for different purposes. The results showed a great diversity of impacts on different areas in the city and classes of users.

Recent research has been looking at the approach to including equity aspects in the design of road pricing systems. Mayeres and Proost (2001) proposed a weighted welfare indicator giving more weight to the benefit/cost imposed on less advantaged groups. The test results showed that road pricing is an important element of the tax reform even with a greater emphasis on equity. Meng and Yang (2002) developed a framework for calculating optimal road toll (to maximise social welfare) with constraints on the spatial equity impact. Jones (2002) proposed a simple approach to address equity concerns through scheme design, exemption, and discount.

While there remain some uncertainties over equity impacts, they mainly relate to issues of scale, which will depend on detailed design, and of design approaches which can be adopted to mitigate these impacts.

2.5.3 Economic impact

There is little evidence on the impact of road pricing on the economy of an area. One reason is due to the difficulty of measuring the impact of transport schemes on the relocation of the business and on change in the economy. There was an attempt to gauge the attitude of the business sector in Singapore after the implementation of the Singapore ALS. However, the result may rather represent the general public view at that time to support the government decision. Later on, there was also an attempt to assess the impact of the road pricing on the economy. Although the general result suggested that road pricing did not generate any adverse impact on economic activity in the city centre, the results are probably not transferable to other cases due to several factors (e.g. parking restriction was also in place, the implementation of the vehicle quota system, the geography of Singapore that limits the competition with other areas, and the rapid growth of the economy in general during that period).

There was also an attempt to evaluate the economic impact of the London congestion charging scheme following claims of a 10% reduction in the retail trade as a result of the charging scheme. However, during the period of the implementation of the scheme there was a disruption of the central line (underground service) which is the major service to the central London, the emergence of SARS, and the start of the war in the Middle East. In addition, the recent report of TfL suggested that the 10% reduction in retail sales was mainly caused by economic depression and that the direct congestion charging effects on retail performance are small (TfL, 2004).

Gerrard (2000) conducted surveys with business in three UK cities (Cambridge, Norwich, and York) on their attitude to the potential impact of the road pricing scheme on the businesses in their cities. The majority anticipated positive impacts due to improvements in the environment and reduced congestion (and the revitalisation of the area), but negative ones on the economy and tourism. More than half of the participants anticipated that the implementation of road pricing may influence their decision on the locations of their businesses.

Clearly, the issue over the impact of the road pricing on the business and economy of the area is still a very ambiguous topic. Further research and empirical evidence are needed to establish the relationship.

2.5.4 Technological aspect of the implementation

In the past, the operation of point-based road pricing schemes was mostly based on manual toll collection or automatic coin collection machines at toll booths. The operation required vehicles to slow down and stop at the toll booth. This system offers a very high level of reliability and enforcement. However, this system creates serious congestion around the toll collection area. Key challenges for technology include reliability, cost of implementation and privacy. In addition, well designed technology can provide greater flexibility in providing for a range of users and vehicle origins, in permitting more complex charging regimes of the kinds outlined below, and in overcoming equity concerns by permitting varying charges and exemptions for different types of user.

In the last decade, there has been a rapid development in charging technology responding to these requirements mentioned above. There are two main streams for the current development of charging technology. The first involves the Dedicated Short Range Communications (DSRC) system. The system is comprised of two main equipments: roadside equipment (RSE) and in-vehicle units (IVU) that enable two-way communication using DSRC. The RSE is connected to a computer which carries out the necessary processing. The system tested in Hong Kong in the 1980s relied on a similar system (Dawson and Catling, 1986). However, the Hong Kong system was claimed to violate privacy, due to its IVU and back-office technology (Borins, 1988). The IVU technology in the early study in Hong Kong was a read-only tag which can only signify the identity of the vehicle to the RSE. The read-only tag cannot convey any information such as credits or charges incurred.

The system in Singapore ERP overcomes this problem by introducing Smart Card technology for use with an IVU (Menon, 2000). Instead of having an account for each vehicle, a smart card contains available funds from which charges are deducted at the charging point. The same system is also being tested and implemented widely in Japan as the basis for the future road user charging system (Kumagai, 2003). The DSRC system operates at free flow level at the charging points. Therefore, it requires a high-level enforcement technology for detecting non-complying vehicles (Blythe and Burden, 1996). The technology currently adopted is the Automatic Number Place Recognition (ANPR) and closed circuit television (CCTV). The ANPR has already been trialled and used effectively in many cases such as the Highway 407 system in Toronto, CityLink scheme in Melbourne, the ERP in Singapore, and recently the ALS in London (Turner, 2001).

The DSRC can operate at different frequencies. This caused a problem in terms of interoperability of different systems developed by different providers (Clark, 2000). An example is the problem in Australia where the toll systems operated in Sydney and Melbourne are based on different standards and are not compatible (Charles, 2001). To overcome the interoperability problem, the Norwegian government set up a company, AutoPASS, to develop and operate the charging technology for tolling facilities in Norway. The new AutoPASS is consistent with both global ISO standards and European standards (CEN). The new specifications are used in the replacement of four systems in Oslo, Trondheim, Rennfast and Hvaler. In addition, currently AutoPASS users can use their cards on almost half of the road toll projects in Norway.

As an alternative to the DSRC based system, the GNSS (Global Navigation Satellite System) and GPRS (General Packet Radio System) can be adopted to operate point and distance charging systems (Catling, 2000). GNSS uses a satellite-based positioning and navigation system to compute the location of a vehicle in a road network. Currently, the USA and Russia provide the two navigation satellite systems (GPS and GLONASS respectively). The EU's Galileo alternative is due to be available in 2008. Since the satellite navigation systems provide only one way communication (from the satellites to the receivers), a cellular phone system is normally used for communication between the vehicle and the control system for the transaction process.

For GPRS, the position can be determined by the data connection via a mobile phone network with an always-on connection. The resulting vehicle positioning system (VPS) allows a more complex charging regime to be implemented. The system also requires minimum infrastructure on the roads. A similar system was tested in Hong Kong and it was proved to be reliable (Catling, 2000). The Swiss and German government also launched the first large scale GPS-based project, which will be soon operable and will charge heavy goods vehicles based on distance travelled (Guillermo Jordan *et al*, 2001).

The key barrier to large scale implementation, especially in an urban area, is the required level of accuracy of the positioning system. At the moment, the accuracy of GPS system is around 10-15 metres whereas Galileo promises to deliver positioning accuracy down to 4 metres. Despite the potential improved accuracy, there exist various black spots in the road network (e.g. tunnels) in which the GNSS may experience some problems. This problem can be overcome by integrating the GNSS with the short range communication system (e.g. communicate with beacons) or a dead-reckoning system (Ochieng, 2003).

2.6 SUMMARY

Road pricing was originally developed from the idea of marginal cost pricing. With a simple concept of imposing the toll equal to the external cost of the trip, the traffic system is believed to achieve its maximum efficiency creating the maximum social welfare for the society. The economic principle of road pricing has been extended to include the analysis with the network equilibrium, improved behavioural responses of users, and appropriate treatment of the supply cost function. One of the main theoretical research areas is related to the concept of 'second-best' optimal tolls. The second-best toll scheme allows for the existence of an untolled alternative route in the network to reflect the imperfect pricing system in the real world. Various researchers have been working on the way to determine the optimal toll under this second-best setting.

A consensus has been reached between academics and practitioners on the merits of road pricing. With this belief of the potential benefit of road pricing, the theory of road pricing has been transformed into a real world policy. Although the original objective of road pricing was related to the enhancement of the economic efficiency of the transport system, various practical objectives have been attached to the implementation of road pricing in the real world including tackling congestion, environmental protection, revenue generation, and urban management. Various charging systems have also been proposed for real world implementation including the point-based charging system, area-based charging system, and continuous charging systems (time, distance, and delay-based charge).

The transition from theory to practice of road pricing can also be observed from the growing number of schemes and studies of the implementation of a road pricing scheme around the world. The key developments have been in the U.K., Singapore, and Norway. Also, there were various discussions and developments of the idea in Hong Kong, the Netherlands, Germany, and Sweden. In the U.K., the key development of road pricing was the implementation of the congestion charging scheme in early 2003 after a long research programme dating back to the 1960s. The scheme in Singapore has been the main case study of the implementation of road pricing for a long time. In a slightly different way, various toll ring schemes are also under operation in Norway aiming mainly to finance the new transport infrastructures.

During the transition phase from the theory to practice of road pricing, a number of questions and obstacles related to the implementation and impact of road pricing have been raised. The main research issues include the public and political acceptability of the scheme, equity impact of the scheme, economic impact of the scheme, and the technology for the scheme implementation.

Overall, from the case studies and the research results progress has been made towards the successful implementation of a road pricing scheme in a wider perspective. Although there are various ways to implement the charge, the point-based charging scheme (particularly the cordon-based scheme) seems to be the favourite option due to its simplicity and readiness of the technology for its implementation. The development of road pricing schemes in practice does seem to concentrate on the acceptability and practicality of the scheme. This may deviate from the original objectives of road pricing. However, it is a real challenge to develop a scheme that is eventually implemented. Thus, achieving the optimal benefit may not be the most important objective of the scheme development. This represents an interesting gap between the theory and practice of the development of a road pricing scheme. In the next chapter, a focus will be made on the analysis of the practical design criteria for a charging cordon scheme.

CHAPTER 3 JUDGMENTAL APPROACH

Out of clutter, find simplicity. From discord, find harmony. In the middle of difficulty, lies opportunity. Albert Einstein

3.1 INTRODUCTION

As mentioned in Chapter 2, the gap between the theory of road pricing and its real world application is significant, due to the issues of public acceptance, technical feasibility, and the cost of implementation (Sharp, 1966; Verhoef *et al*, 1995; Stenman and Sterner, 1998). Different charging regimes have been developed and studied including time-based or delay-based charging, distance-based charging, cordon or boundary-based charging, and area based charging. May and Milne (2000) used modelling case studies to compare the performances of different road pricing regimes. They showed that the cordon based scheme is the least optimal of the charging regimes. However, the review in the previous chapter shows the popularity of cordon based road pricing. This is believed to be due to the practicality and ease of use of the charging cordons.

This chapter aims to investigate the way in which transport planners (practitioners) design a road pricing scheme. In particular, the cordon charging system is the core of the study in this chapter. The term *"judgmental cordon design"* is used to describe the process to identify the best locations to levy the charges and to specify the optimal charge levels based on professional judgment.

Section 3.2 reviews the literature to identify design criteria; then surveys with six UK local authorities are discussed in Section 3.3. Section 3.4 presents the results of the surveys and finally Section 3.5 discusses the results, and draws conclusions.

3.2 REVIEW OF THE PRACTICAL DESIGN CRITERIA

The Smeed Report (Ministry of Transport, 1964) identified nine criteria for the design of congestion charging systems. These criteria can be used to help determine whether the scheme design is likely to perform effectively and to be feasible. The criteria include:

(1) charges should be closely related to the amount of use made of the road;
(2) it should be possible to vary prices for different areas, time of day, week, or year and classes of vehicle;
(3) prices should be stable and readily ascertainable by road users before they embark upon a journey;
(4) payment in advance should be possible although credit facilities may also be permissible;
(5) the incidence of the system upon individual road users should be accepted as fair;
(6) the method should be simple for road users to understand;
(7) any equipment should possess a high degree of reliability;
(8) it should be reasonably free from the possibility of fraud and evasion, both deliberate and unintentional;
(9) it should be capable of being applied, if necessary, to the whole country and to a vehicle population expected to rise to over 30 million.

These criteria were set over 40 years ago, and they still remain valid over time. However, other criteria which have emerged since include:

(10) the system should allow occasional users and visitors to be equipped rapidly and at low cost;
(11) the charge recording system should be designed both to protect individual users' privacy and to enable them to check the balance in their account and the validity of the charges levied;
(12) the system should facilitate integration with other technologies, and particularly those associated with driver information systems (May, 1992).

Criteria 1 and 2 are concerned with the ability to levy an optimal charge; criteria 3, 4, 6 and 7 concern the ability of drivers to respond in an optimal way to the charge levied; criteria 8 and 9 and to some extent 7 address the efficient operational of charging regime in terms of implementation costs (9), operation (7), and enforcement (8); criteria 10, 11 and 12 affect the efficiency and success of the operation. Some of these criteria are more related to technological issues which are out of the scope of this research, i.e. criteria 4, 7, 8, 9, 10, 11, and 12. Most of these criteria (1- 9) aim to ensure that the design meets the three important factors of the congestion charging scheme design including:

(i) Effectiveness of the scheme
(ii) Public acceptance of the system
(iii) Practicality of the implementation.

In the following sections, the design approach adopted to achieve these three main targets is discussed.

3.2.1 Design options

The focus of this chapter is the cordon based charging system. Thus, the design options can be narrowed down to the three main systems suggested by various literature (Neuenschwander, 2000; Shepherd *et al*, 2001). These are:

Single Road or Motorway Charging: The idea for charging on the major road infrastructure was first suggested by the French engineer, Depuit (1844). The main advantage of charging on individual roads is to minimise the affected groups. There are several examples where urban road pricing is imposed on single roads including Marseille (tunnel) and San Francisco-Oakland Bay Bridge (Dittmar *et al*, 1994; Neuenschwander, 2000). Most of the tolls are on new major road facilities, i.e. bridges, tunnels, or motorways. The major barrier of this system is the lack of alternative routes, which can lead to public acceptance problems.

Simple Cordon Charging: Only one charging cordon is used, normally where a *'natural position'* for the cordon already exists for example a ring road, river, canal, coast or rail track (Neuenschwander, 2000). The charge level is usually defined as a single charge for all crossing points to simplify the system.

Complex Cordon Charging: The complex cordon system can be viewed as a development of the simple charging cordon, where additional cordons or screen lines are added to the charging cordon system. There are various reasons to move to a more complex system. Shepherd *et al* (2001) mentioned the possible objectives of using additional cordons as follows:

- tackling congestion outside the first cordon;
- extending the area of influence of road pricing;
- controlling traffic to the inner city area outside the main centre;
- relating charges more directly to distance travelled;
- reducing the charge at any one crossing point, and hence the boundary effects.

Shepherd *et al* (2001) mentioned the reasons for adopting screen lines as follows:

- to control orbital traffic generally;
- to control access to particular high traffic generators;
- to protect a specific bottleneck or other source of congestion.

Dawson and Brown (1985) and May *et al* (1996) showed that a more complex cordon system could achieve higher benefits when compared to a simpler system as a result of these factors.

3.2.2 Dealing with the objectives of the scheme

May (1992) stated that the design of a congestion charging system depends heavily on the objectives of the scheme. This can be confirmed by the differences in the design of the Singapore Area Licensing Scheme (ALS) (Holland and Watson, 1978) and its modification later on as explained in Chapter 2; the Hong Kong Electronic Road Pricing (ERP) pilot scheme (Harrison *et al*, 1986); the London congestion charging scheme (May, 1975); and the Oslo toll ring (Larsen and Ramjerdi, 1991). As described in Chapter 2, the Oslo toll ring aimed primarily at raising revenue for financing new roads and it was designed to change the traffic pattern as little as possible. The area coverage of the Norwegian cordon is wider than other cases and located on the trunk road system rather than the urban road network. Norwegian toll rings apply a low charge level which will not reduce traffic demand whereas in the other cases the charge level is higher.

The Hong Kong pilot system planned to use a more complicated cordon system where the main objective was to change travel behaviour to more efficient patterns by enhancing the use of public transport and time choice (Harrison, 1986). The structure of the cordon system proposed in the Hong Kong case involved using a number of cordons with a different charge level on each cordon, and the charge also varied by time of day. The cordons and screen lines are located in the most congested area. The Singapore ALS and the London congestion charging scheme adopted a simpler cordon, aimed at reducing the congestion in the core area of the city by using a single cordon around the city centre and simple charge structure. The cordon was located inside the inner ring road of the city in order to provide the diversion route for through traffic. However, it is noteworthy that the current system in Singapore is being modified by providing an incomplete second cordon. This represents the evolution of the initial scheme to a more complex one.

3.2.3 Dealing with public acceptability

Equity, fairness, and simplicity

As reviewed in Section 2.5.2, there are two main dimensions of equity: horizontal and vertical equity. The actual design of a charging cordon or road pricing scheme can influence the equity impact to some extent. For the vertical dimension, exemption or discount schemes (as adopted in the London congestion charging scheme) can be an approach to alleviate the equity impact. The location or structure of the charging system itself may have some small impact on vertical equity; for example, the planner may not want to locate the charging cordon if one area with low-income population will be the most affected group.

For the horizontal dimension (spatial equity) there are various examples of the attempt to minimise the impact of the road pricing scheme on the public. Holland and Watson (1978) stated strongly that the cordon should not split the business area, since this will give advantages to some business places or shops just outside the cordon.

From a slightly different angle, a fair road pricing system can be interpreted as being a system which charges people as closely as possible to their contribution to the congestion or environmental problem (Jones, 1998). Using additional cordons or screen lines can enhance the fairness of the cordon by imposing charges better related to length of a trip (Holland and Watson, 1978; May, 1992; Shepherd *et al*, 2001).

However, Sheldon *et al* (1993) stated that a more complex cordon system would experience difficulties in implementation due to the opposition of the public. A simple system tends to be more acceptable. Politically, Rom (1994) also suggested that a congestion charging programme which did not rely on complex strategies of implementation would be more politically attractive. Where it is too difficult to adopt a more complex cordon system, it is appropriate to concentrate the charge only on the central area of the city, considered as being the most congested part of the city. This design can be seen in the aforementioned cases, i.e. the Singapore ALS and ERP, and the London congestion charging scheme.

Jaensirisak (2004), based on a stated preference survey, suggested that acceptable road user charging schemes can be designed by limiting the area of charge to within the city centre and having a fixed charge per day. Support would be increased significantly if the scheme was promised to bring substantial environmental improvement.

Freedom of travel

The implementation of a congestion charging scheme will cause infringements on freedom of travel. The key element to improve public acceptance is the provision of travel alternatives in terms of both alternative routes and modes. The combination of public transport improvement and congestion charging can be viewed as an element of an integrated transport strategy that can reduce congestion and also enhance public acceptance (Goodwin, 1989).

The design of the Singapore ALS boundary provided the ring road as a diversion route for through traffic (Holland and Watson, 1978). In the Hong Kong case, the congested corridors with good public transport service were considered as good candidates to be tolled (Transpotech, 1985). When the objective is to raise revenues, the provision of diversion routes is not appropriate; instead a lower charge level is adopted to reduce the public opposition (Larsen and Ramjerdi, 1991; Lewis, 1993).

Alternative travel periods or days should also be considered; for example in Bergen, the tolling system is not under operation on Saturday and Sunday which is aimed to provide the opportunity to people wishing to carry out activities in the central area with the free use of the road network. The Singapore ERP also allows free driving at weekends (Menon, 2000), and this has had a significant effect on the acceptance level of the scheme (Lewis, 1993; Langmyhr, 1997). The congestion charging scheme in London is also operated between 07:00 – 18:30 during the week days.

3.2.4 Dealing with adverse impacts

Economic and land use impact

As discussed in the previous chapter, the impact of road pricing on the economy and land use is a serious issue, and it is extremely difficult to isolate the effect of any one element of transport policy on land use (May, 1992). There are several suggestions on how to minimise the impacts of road pricing on the economic and land use changes:

- Locating the charging boundary between land use types. In practice it may be appropriate to use existing geographical boundaries such as rivers, railways, canals, or mountains as the boundary of the charged area (Shepherd *et al*, 2001). However, in some cases where these physical barriers are not appropriate, boundaries of land uses or other boundary alternatives should be considered, such as ring roads;
- Trying to minimise the interruption of trips serving basic needs and residential areas, e.g. trips to school, hospital, or public services. This implies that it is better to charge business trips rather than residential trips and other trips serving basic needs (e.g. charging inbound traffic to a CBD). By doing this, the land use change can be kept to a minimum and the economy of the area can be maintained (Holland and Watson, 1978).

Dispersion of congestion and environmental impacts

There is a danger that current traffic using the tolled roads could be diverted to other areas, causing congestion and pollution problems in those areas. Traffic could also be diverted to untolled periods. Evidence in the Singapore ALS showed dispersion of traffic to the period prior to the charging period and the period after the charging time. There are two facets of this problem:

- the first is that we can view this as a good effect as the concentration of traffic is spread over different parts of the network or different time periods, reducing the congestion and the pollution problem in the peak period;

- the other is that the charge spreads the problem over a longer time period and a wider city area; the effect is not only an increasing pollution and congestion problem but also an increasing safety problem particularly when traffic diverts to residential areas or local streets which are not designed to cope with it.

Dispersion over space

In the case where the main objective is to reduce pollution over a wide area of the city (not only just the area inside the cordon), the main idea is to suppress trips rather than divert them to other areas. This could be achieved by designing the cordon to capture most of the traffic without providing diversionary routes. On the other hand, if the main objective is to increase efficiency and reduce congestion in a specified area, e.g. the city centre, the diversion of through traffic can be allowed (Holland and Watson, 1978). In this case, the cordon location is normally placed just inside the ring road (Shepherd *et al*, 2001). The capacity of the diversion route also has to be compared with the potential diverted traffic. Another design approach to tackle the dispersion problem is to use a more complex cordon. A number of screen lines and cordons are used with different charge levels on different cordons and screen lines. This finer charging system is expected to smooth traffic at the boundary of the cordon (Holland and Watson, 1978).

Dispersion over time periods

In the Singapore ALS, the initial restriction period was designed to operate from 7.30 am to 9.30 am in order to reduce congestion in the morning peak. After implementation, congestion developed after 9.30 am and the problem was solved by extending the restriction period to 10.15 am (Holland and Watson, 1978). A different approach was proposed to prevent the same problem in the Hong Kong case where the charge structure was designed to vary by time period (Harrison, 1986). The case of the London congestion charging scheme showed another approach where the flat rate charge is designed to cover the a longer period (7am – 6.30pm). The scheme objective distinguishes the case of Norwegian toll rings from the other cases, since the problem of the dispersion of the traffic over the time period seems not to be a problem where a low charge is applied over the whole day. More specifically, dispersion was to be avoided to maintain revenue.

3.3 ISSUES ON DESIGN CRITERIA AND SURVEY DESIGN

Table 3-1 below summarises the key design criteria of a charging cordon scheme from the review in the previous section. These criteria will be tested by a survey with practitioners.

Design criteria to avoid the adverse impacts
▪ The design should ensure the provision of sufficient alternative routes for drivers who want to bypass the charge area.
▪ The design should avoid the dispersion of the environmental or congestion problem to other areas.
▪ The cordon should cover only areas having good public transport service.
▪ The design should leave the facilities for interchange outside the cordon (e.g. park and ride or car park).
▪ The design should ensure that all entry points to the charge area are either charged or closed.
▪ The design of the entry points should not be visually unattractive.
▪ The design should place cordons at boundaries between land use types.

Design criteria to gain public acceptance
▪ The cordon structure should be simple and easy to understand.
▪ The charge structure should also be simple and easy to understand.
▪ The charge should be at a level which is acceptable to the public.
▪ The charge should be perceived as fair by the public.
▪ The design should avoid the problem of local inequities (e.g. people just outside the cordon needing to access places just inside)
▪ The design should avoid the problem of commercial inequities (e.g. with the same type of business, one is just inside the cordon and the other is just outside the cordon)
▪ The design should aim at charging the traffic which contributes most to congestion and pollution.
▪ The design should aim at charging the traffic which is of least benefit to the area.
▪ The design should avoid charging the city's residents.
▪ The design should avoid charging people from low income area of the city.

Design criteria to ensure practicality
▪ The number of charging points should be minimised to reduce capital costs.
▪ The system should be designed to limit the scheme's operating costs.
▪ The design should avoid types of road that cannot be tolled e.g. motorways.
▪ The design should avoid areas or locations that may cause technological or communication problems to the system.
▪ The cordon should be located wholly inside the city authority area.

Table 3-1 Possible design criteria for practical a road pricing scheme design based on literature review

The research questions to be investigated, by questionnaires and in-depth interviews, can be defined as follows:

i.) *What are the objectives of the congestion charging scheme considered by the local authority and to what extent do these objectives influence the design of the charging cordon compared to other factors?*

ii.) *Is the simplicity aspect of the charging system a necessary design criterion for a practical cordon, and to what extent can the single charging cordon system be modified by introducing additional cordons or screen lines to give the possibility of higher benefit of the scheme?*

iii.) *Are the design criteria found from the literature consistent with the opinion of the respondents in the survey, and to what extent do these design criteria influence the design of practical cordons?*

iv.) *Are there any other necessary conditions that the cordon design must follow?*

A two stage survey was designed. In the first stage, questionnaires (see Appendix A) were sent to the respondents and the answers from the questionnaires were analysed. The second stage of the survey involved using an in-depth interview to probe the points raised from the answers from the questionnaires.

Question one starts the questionnaire by asking the importance of each objective and then question two asks whether there are any differences in the design of the cordon to meet different objectives. The answers to question one and two can reveal the set of objectives considered by local authorities and the way local authorities try to design the scheme to meet these objectives.

Question three asks about the general design of the charging cordon in each city in terms of the cordon and charge structure. Question four asks the respondents about the possibility of using additional cordons or screen lines. These two questions are used to investigate the general design of the cordon and the possible level of complexity of the design, i.e. cordon and charge structure.

Question five asks the respondents to express their opinions on whether each design criterion found in the literature review should be considered in the design of the charging cordon. The respondents are also asked to address any other criteria apart from those provided.

Question six asks the respondents which conflicts arise between the charging cordon design to meet the objective and the constraints they may have experienced. The answers to this question can be used to find whether each local authority gives more priority to public acceptance and adverse impact issues compared to the objectives of the scheme in the design of charging cordon.

Question seven is designed to ask about the objective of raising revenue, which was strictly excluded from acceptable objectives of congestion charging in the U.K. legislation. The answer to this question can show whether raising revenue is one of the objectives of the scheme.

The next stage of the survey was the in-depth interview. The structure of the interview was designed to be a semi-structured interview where there were three main discussion topics including the characteristics of the case, the objectives of the scheme, and the design process of the charging cordon. Interesting points from the questionnaires or from the discussion during the interview are probed into the detail.

3.4 SURVEY RESULTS

3.4.1 Responses to the questionnaires

The summary of the responses to some of the questions is shown in Appendix C. In this section, we will analyse the answers in order to answer the research questions set earlier.

Objective of the scheme and hierarchy of objectives

Table 3-2 above shows the answers to this question from the survey. Despite the absence of the explicit option of an objective of revenue generation in the questionnaire, some local authorities raised this objective in their responses as well as the objective of economic regeneration. The answers show the wide range of objectives of using congestion charging, with most of the schemes being expected to serve more than one objective. Nevertheless, the main objectives of most of the local authorities in this survey are to reduce congestion and increase efficiency.

Objective	Bristol	Birmingham	Edinburgh	Leeds	Manchester
Reducing congestion	H	M	H	H	M
Environmental protection	N	L	M	M	M
City centre management	H	N	M	M	M
Increase efficiency	H	L	H	M	M
Redressing inequity in transport system	N/A	L	N/A	M	M
Raising revenue (LAs added this objective by themselves)	N/A	H	N/A	L	N/A
Economic regeneration	N/A	N/A	R	N/A	H

Table 3-2 Response to question regarding the objective of the scheme[6]

Conflict between the different objectives

The responses to this question tend not to show great concern that different designs are needed to meet different objectives.

General design of charging cordon

It is politically sensitive to give the detail of the design of each case here. Instead, the general response will be given and discussed. Four out of the five local authorities in the survey have some idea about the design of the actual cordon. From these four local authorities, the general responses to question three asking about the cordon location and charge structure are as follows:

- the cordon will cover the core or city centre area of the city;
- the cordon will use a ring road as the reference boundary (normally the cordon will be situated inside the inner ring road);
- the size of the scheme will be relatively small compared to the whole area of the city;
- three out of four cases only consider a single cordon application; the other case considers different options including a city centre cordon, double cordon (inner

---

[6] H = high, M = medium, L = low, N = not an objective, R = raised but not specify level, N/A = no response

and outer ring road), and a wide area cordon (outer cordon with area licensing scheme instead);

- the charge structure will be very simple, with a uniform charge for the whole day or two charge levels for peak and off-peak;
- three out of four cases expressed that the charge would be operated only in the peak period and for inbound traffic only;
- the charge level would be a uniform rate for all toll points on the cordon;
- three out of four cases are considering the application of an area licensing scheme or charge only once a day regardless of the number of crossings.

Additional cordons and screen lines

Question four particularly asked further about the extent to which the complexity of the cordon structure could be reconsidered. Only one out of four local authorities accepted the possibility of introducing additional cordons and screen lines in the future. It is fair to conclude that the simple cordon and charge structure is an important element of the design of a practical cordon.

Design criteria

Table 3-3 - Table 3-5 show the responses to this question in the survey. It shows that most of the design issues, found during the literature review, are consistent with the views of the respondents in the survey. Eight design issues out of 22 receive a unanimous positive answer from all respondents. The design issues related to practicality receive the most negative responses. It is noteworthy that the understanding of the respondents to this issue is still at an early stage, since most of the schemes are still in the early design stage. Two issues in the group of public acceptance receive strong negative responses including:

- the issue of whether the design should avoid low-income areas and
- whether the design should avoid a charge on residents (only Bristol agreed with this issue).

Also two issues received a vary of responses which are:

- whether the cordon should be placed between land-use types (Bristol and Leeds disagree, others agree); and

- whether the design should aim at charging the traffic which is of least benefit to the area (Edinburgh and Manchester agree, others disagree).

The response from Edinburgh is significantly different from others especially when compared with the group of Bristol, Birmingham, and Leeds. The design issues of provision of a bypass route from the charging area and public transport quality in the charging area received a negative response from Edinburgh whereas other local authorities agreed with these issues.

Avoid the adverse impacts	**Bristol**	**Birmingham**	**Edinburgh**	**Leeds**	**Manchester**
• The design should ensure the provision of sufficient alternative routes for drivers who want to bypass the charge area.	A	SA	D	SA	SA
• The design should avoid the dispersion of environmental or congestion problem to other areas.	A	SA	SA	A	A
• The cordon should cover only the area having good public transport service.	A	A	D	A	SA
• The design should leave the facilities for interchange outside the cordon (e.g. park and ride or parking facility).	A	A	SA	SA	A
• The design should ensure that all entry points to the charge area are charged or closed.	SA	SA	SA	SA	A
• The design of the entry points should not be visually unattractive.	A	A	SA	A	SA
• The design should place cordons at boundaries between land use types.	D	SA	A	NC	A

Table 3-3 Responses to questionnaires regarding the design issue related to adverse impacts of the scheme[7]

Conflict between objectives and design criteria

Similarly to the response to question three, there are no very interesting answers to this question from the respondents. Most of the answers tend to say that there is no conflict between the design to meet the objectives and other constraints in their cases. This point is investigated further in the in-depth interviews.

---

[7] SA = Strongly agree, A = Agree, NC = Not consider, D = Disagree, and SD = Strongly disagree (N/A = no response to the question)

Practicality	Bristol	Birmingham	Edinburgh	Leeds	Manchester
- The number of charging points should be minimised to reduce capital costs.	N/A	A	D	A	D
- The system should be designed to limit the schemes operating costs.	N/A	A	D	A	A
- The design should avoid the types of road that cannot be tolled.	N/A	N/A	N/A	A	N/A
- The design should avoid areas or locations that may cause technological or communication problems the system.	N/A	NC	N/A	A	NC
- The cordon should be located wholly inside the city authority area.	N/A	NC	SA	SA	SD

Table 3-4 Responses to questionnaires regarding the design issue related to practicality of the scheme

Gain public acceptance	Bristol	Birmingham	Edinburgh	Leeds	Manchester
- The cordon structure should be simple and easy to understand.	SA	A	SA	SA	SA
- The charge structure should also be simple and easy to understand.	A	A	SA	SA	SA
- The charge should be at a level which is acceptable to the public.	A	A	A	A	SA
- The charge should be perceived as fair by the public.	A	A	A	A	SA
- The design should avoid the problem of local inequities (e.g. people just outside the cordon needing to access places just inside)	A	A	NC	A	SA
- The design should avoid the problem of commercial inequities (e.g. with the same type of business, one is just inside the cordon and the other is just outside the cordon)	A	A	NC	A	SA
- The design should aim at charging the traffic which contributes most to congestion and pollution.	SA	A	SA	A	A
- The design should aim at charging the traffic which is of least benefit to the area.	N/A	D	SA	NC	SA
- The design should avoid charging the city's residents.	A	SD	D	NC	NC
- The design should avoid charging people from the low income area of the city.	N/A	D	NC	NC	D

Table 3-5 Responses to questionnaires regarding the design issue related to public acceptability of the scheme

Spending revenues

The local authorities have a very clear idea about the way to spend the revenues from the scheme. Most of the generated revenues will be used to invest in improving the public transport service.

3.4.2 In-depth interview analysis

The detailed analysis of the interviews can be found in Sumalee (2001). A brief summary of the finding in each case is presented in this section. The background for each case can also be found in Appendix B. Note that the information provided in this section is based on the interviews conducted in 2001.

Birmingham and West Midlands area

The process of cordon design starts by looking at a possible single cordon around the centre of each city. Only the inner ring roads are considered as the potential boundaries of charging cordons. Birmingham, Walsall, and Solihull have feasible cordon boundaries based on the characteristics of the inner ring roads of these cities whereas the other cities have problems with their ring roads. From the interviews, the reasons why the ring roads in other cities are not feasible or satisfactory to be used as the cordon boundary are as follows:

- The ring road is too small in terms of the surrounded area, too small a cordon may not be able to reduce the congestion sufficiently and generated revenues from the scheme could be too low;
- The ring road is not completed or in some cases there is not a clear ring road at all;
- The ring road does not provide a clear boundary between the business area and residential area.

It is also less politically feasible to implement the scheme in Solihull or Walsall, which have a feasible cordon boundary, when the more congested cities like Coventry or Wolverhampton will not implement a charging scheme due to the lack of a feasible cordon.

Considering only Birmingham city, the potential cordon would be situated just inside the middle ring road. The walking distance from the boundary of the cordon to city centre is about ten to fifteen minutes. There is a plan to increase the yellow line parking control around the cordon in order to prevent the dispersion of parking demand at the fringe of the cordon. The design would also plan to allow free lanes on the main motorway passing through the city, since the ring road is predicted not to be able to cope with diverted traffic from this main through route. The charge level will be low in order to maintain support from the public and the charge will be a uniform rate all day.

Bristol

Bristol considers a small cordon covering the city centre, and the cordon would be located just inside the inner ring road covering most of the inner parking zones of the city. The underlying reason for the small scheme is political and public acceptability, since the group affected by the scheme can be kept to a minimum and this small scheme seems to be able to express the objective of the scheme clearly to public. The design of the cordon location would also try to minimise the number of crossing points or toll points by using the river on the south of the city to form the boundary.

The recommended scheme has 14 entry points and includes the city's main Broadmead shopping area, the Centre, West End, and Harbourside but excludes the main inner city residential area. The bus station is also deliberately left just outside the cordon. There are a number of car parks outside the charging boundary. It was indicated that for Bristol there is a potential extension of the scheme toward a more complex system but this possibility really depends on the success of the forthcoming scheme whereas in this stage it is only possible to implement the single cordon with a simple charge structure. The charge will be in the morning peak and for inbound traffic only and the level of charge will be at a low level. It is accepted that this low charge level may not be able to reduce congestion significantly but can be more politically acceptable.

Durham

The design of the charging system in Durham is different from other cases. Only one toll point has been located at the access point to the peninsula area from the market place. The geography of the peninsula area forms a natural closed cordon. The charge level is £2 and operates from Monday to Saturday from 10 am – 4 pm. The design of the scheme tries to avoid the effect on necessary trips, such as school trips, by providing an exemption to the parents of the students under 6 years old studying at the schools in the peninsula area. The scheme should not affect the employees in the area since normally most of them do not use a car for business trips due to the lack of parking space.

Edinburgh

The inner ring road and outer ring road are considered the possible boundaries for charging cordons. The implementation of two cordons, i.e. inner and outer cordon, may cause problems in the doughnut area in between these two cordons in which there may be problems of increased traffic, and the public transport services in the orbital direction around the city centre are not at a good level. One of the points mentioned about the design of this outer cordon is that the cordon must be wholly inside the city bypass and it has to be purely inside the City of Edinburgh Council's area. The location of the inner cordon should cover the inner parking zone, but not as far as the boundary of the outer parking zone which in this design aims to reduce the parking dispersion problem around the cordon boundary.

It was mentioned that it becomes inevitable to include one of the residential areas, which is just outside the boundary of the initial cordon, inside the cordon in order to avoid worsening the congestion problem in that already congested area. The inner cordon also uses some of the physical boundaries such as parks, railway line, and river to form the closed charging boundary to help minimise the number of toll points and make the cordon easily recognised. At each toll point, there will be an escape route for those who want to avoid the charge by diverting to other routes, but the diversion route from one side to the other side of the cordon will not be advised to drivers in order to discourage rat-running traffic.

In the case the twin cordon system is implemented, the charge will be higher on crossing the inner cordon and lower on crossing the outer cordon whereas the single uniform charge will be used if only the inner cordon is implemented. At the time of the interview, the level of charge had not definitely been decided but it would be at an acceptable level to the public and politicians.

Leeds

The system considered in Leeds is an area licensing scheme (ALS). Similarly to other cases, Leeds considers the inner ring road and outer ring as good candidates for the location of charging cordons. However, the outer ring road cordon is considered to cause a major political acceptance problem because many people living inside the cordon would be charged. Therefore, the inner ring is chosen as the reference boundary of the charging cordon subject to the completion of the inner ring road.

The city council has conducted several modelling tests on whether the capacity of the inner ring road can cope with the diverted traffic, and it is concluded that the capacity of the inner ring road is sufficient. It is accepted that the inner cordon does cover the whole congested area in the city centre but it does not address the congestion problem of the whole city. The detailed design of the cordon is trying to minimise the number of toll points, trying to include the University area inside the cordon, and to leave the major hospital outside the cordon.

Politically, the hospital inside the central area must be outside the cordon although it is located just inside the inner ring road. This political requirement causes a conflict with the requirement to include the university area inside the cordon where this area is actually just outside the cordon. The charge level has been suggested based on the requirement that it should reduce the congestion in the city centre, generate a significant amount of revenue, and receive support from public and politicians. According to the concept of ALS, the charge structure will be a uniform charge throughout the day. Similarly to the Bristol case, there is a possibility to introduce additional cordons in the future in which the second and third cordon may be located just inside the outer ring road and the middle ring road respectively in which the charging system would change from ALS to a more complex cordon charging structure.

Manchester

Manchester is still in a very early stage of the plan to implement a possible congestion-charging scheme. From the geography of Manchester city, it is too difficult to find an appropriate cordon that is wholly inside the Manchester City Council's area, and it may cause a problem to the economic development of Manchester if the charging scheme is introduced in Manchester alone. It is thus necessary that all of the local authorities in Greater Manchester come to an agreement in introducing a congestion charging scheme. It is affirmed that a simple charging system that is just good enough to make the scheme work would be preferable to a more complicated system that has to rely on the high technology.

3.5 FINDING FROM THE SURVEY

The questions about differences in design to meet different objectives and about the conflict of the design to meet objectives and constraints are not well answered, since most of the answers do not express concern over these issues. The discussion with local authorities in the in-depth interviews shows that the issues of public acceptance and adverse impact are more important to the design compared to the objectives of the scheme.

During the discussion, the design of the cordon is regularly associated with public acceptance and adverse impact issues whereas the objective of the scheme is only mentioned when it is to be decided whether the scheme is worthwhile. It was regularly mentioned that the objectives of the scheme, particularly the objective to reduce congestion, cannot be met by only using congestion charging. In fact, it is expected that the charging cordon alone will not be able to reduce the congestion significantly but the revenues generated from the scheme are the key to success. The revenue generated from the scheme will be invested in improving the public transport service.

The other point found in the responses to the question about the cordon design is that the design of cordon location and structure is a separate process from the decision on the charge level. The charge level may be even defined in advance at an acceptable level before the location of the cordon is decided.

It is found from the questionnaire and in-depth interview that the design criteria found in the literature review are not all necessary. Some of these criteria are "hard constraints" which the cordon design must follow, but some are not. We refer to the other type of constraints as "soft constraints". Some criteria are not even agreed by most of the respondents, such as the issue to avoid charging residents of poor income areas.

It seems to be that the design tries to find the cordon that strictly meets the set of the hard constraints which could be considered as an "acceptable design" which can be acceptable and will not cause the new problems. The common features or hard constraints of the "good enough cordon" are as follows:

- Use the ring road as the skeleton design of the cordon and as the diversion route;
- The cordon and diversion route must be wholly inside the authorised area of local authority;
- However, in the case that there is no obvious ring road, try to find a road network that can form the escape routes from the toll points instead;
- The capacity of the diversion route should be able to cope with the diverted traffic;
- Ensure that the problem of the dispersion of congestion to the surrounding area around the cordon is kept to a minimum;
- Cordons should concentrate on the central area of the city, even if congestion also exists elsewhere in the city;
- Include major trip attractions but exclude sensitive locations if possible, i.e. hospitals or bus or train stations;
- Associate the cordon boundary with the controlled parking zone which can give a buffer area to protect the dispersion of parking demand around the fringe of cordon;
- Minimise the crossing points by using a natural boundary, such as park, river, or railway line;
- Keep the boundary as simple and clear as possible and try to use key landmarks or natural boundaries;

- Use a simple charge structure that is easy to understand and remember;
- Charge level should be defined at an acceptable level to the public;
- Finally, the scheme must be worthwhile to implement in terms of the benefit of the scheme given the possible cordon design and charge level.

Several issues are discarded from this list, e.g. separation between land use types, fairness and equity issues, residential or poor income areas. These are the "soft constraints".

A simple charging cordon is clearly preferable to a complex one. However, the interviews show that local authorities are also aware of the greater benefit that they could achieve by using a more complex charging system. Most of the comments on this issue are of the kind: *"at this stage, we are trying to find the system that is just good enough to make this scheme work and start rather than trying to find the optimal design that may not be possible to implement, but of course there is a possibility to extend the system toward the more complex system subject to the success of this starting scheme"*. From this statement, it is clear that the practical design of the cordon is trying to find a scheme that could be implemented now (an "acceptable scheme") and leave the space for further development or evolution of the scheme in the future towards a more effective and optimal system.

3.6 SUMMARY

Following the objectives of the research, an extensive review of the literature regarding the design of charging cordon systems has been carried out. The cordon design in different studies including Singapore, Hong Kong, London, and the Norwegian toll ring reveals the design process of charging cordons and issues that should be considered. The main conclusion from the review is the set of design issues found and three factors that the design tries to achieve including effectiveness, acceptability, and practicality of the scheme. The research investigated these issues further in the context of the UK, and six local authorities in the DfT Congestion Charging Development Partnership (Birmingham, Bristol, Durham, Edinburgh, Leeds, and Manchester) participated in the survey carried out in this research. They were asked to complete questionnaires and participate in the in-depth interview.

The results from the survey show that the design of practical cordons concentrates mainly on the issues of public acceptance and adverse impacts. The effectiveness of the scheme can be achieved by using the revenues generated from the scheme instead of using the charging cordon directly to tackle the problem. Some of the design issues found in the literature were found not to be necessary as criteria for cordon design. Local authorities in the survey consider a smaller set of design criteria that will ensure the scheme would be able to be implemented. These criteria are referred to as the "hard constraints" of the cordon design (see Section 3.5). Thus, most of the designs will start with a simple scheme (despite probably achieving less benefit), with the possibility of extension of the scheme subject to the success of the starting scheme.

The design of practical cordons puts more emphasis on the issues of public acceptance and adverse impacts which are considered as sensitive issues to public and politicians. It is clear that this practical design is not an optimal design to meet the objective of the scheme. The purpose of this research is to combine these practical design constraints with a mathematical model that can produce an optimal design in which the scheme is still able to be implemented in the practice (this is dealt within Chapter 6 and Chapter 7).

CHAPTER 4 FORMULATION OF OPTIMAL TOLL DESIGN PROBLEM

Model should be used not believed. Henri Theil

4.1 INTRODUCTION

In Chapter 2 the review of the development of road pricing reveals two parallel streams of the evolution of the research and practice for this transport instrument. The first stream, as discussed in Section 2.2 Chapter 2, involves the theoretical construction of the optimal road pricing policy from an economic perspective. On the other hand, the development presented in Section 2.4 2.3in Chapter 2 demonstrates a different perspective on how this policy is put into practice. In Chapter 3 the judgmental design approach for road pricing schemes is discussed. The judgmental approach is based mainly on the practical aspect of the scheme while little attention is allocated to the issue of optimality and/or performance of the scheme. The furthest development in terms of scheme design is to exploit transport modelling as an evaluation tool rather than as a designing tool. This chapter concentrates on using a transport model as a design tool rather than just as an evaluation tool.

In this chapter, the problem of optimal road pricing design is formulated mathematically as an optimisation problem. The key and problematic aspect of the optimal road pricing design problem is the travellers' responses to the toll imposed. Naturally, if one tries to evaluate the real benefit of the road pricing scheme, one needs to allow the travellers to respond to the tolls imposed (by changing routes, modes, or deciding not to travel) before calculating the benefit of the scheme. Obviously, a modelling concept needs to be adopted to represent this behaviour. One of the most famous modelling philosophies for representing travellers' decisions in a transport network is the concept of Wardrop's user equilibrium (UE). The next section defines the concept of Wardrop's user equilibrium (Wardrop, 1952). There are various ways to mathematically form the UE condition including equivalent optimisation (Beckmann *et al*, 1956), variational inequality (Smith, 1979; Dafermos, 1980), nonlinear complementarity (Aashtiani and

Magnanti, 1981), and gap function (Hearn *et al*, 1984). The algorithms for solving the optimal road pricing design problem discussed in Chapter 5 are closely linked with the way the UE condition is formulated. These different UE formulations will be explained in the next section as the background for the review of different algorithms in Chapter 5.

By including the UE condition as one of the constraints of the optimal road pricing design, the problem becomes one of the most complex optimisation problems namely a Mathematical Program with Equilibrium Constraints (MPEC). The structure of MPEC can also be considered as a Stackelberg game where there are two types of players, the leader and followers. If the equilibrium model can be reformulated as an optimisation problem (UE in an equivalent optimisation form), the MPEC becomes a bilevel optimisation programming problem (BLPP). Both MPEC and BLPP have been attracting attention from various researchers in different disciplines due to their importance and relevance to the real world problems and the theoretical challenge in solving them. The next chapter will be devoted to reviewing the algorithm development for MPEC and BLPP (in particular for the optimal road pricing design problem). In this chapter, the main focus is to explain the formulation of the optimal road pricing design problem as an MPEC (also equivalently as Stackelberg game and BLPP) in Section 4.3. In addition, different instances of optimal road pricing design will also be discussed.

In Section 4.3, the objective functions of the optimal road pricing design problem are not exactly defined. Section 4.4 discusses the mathematical formulations of the well-known social welfare function and a measure of the equity impact (using the concept of *Gini* coefficient). Finally Section 4.5 summarises the chapter.

4.2 USER'S EQUILIBRIUM CONDITION

4.2.1 Notations

The traffic network is represented by a graph $G(N,A)$ where N is a set of nodes and A is a set of arcs or links. A link $a \in A$ is defined by two nodes $i,j \in N$ where $i \neq j$; i is the starting node and j is the ending node of link a. A subset of nodes are origin and destination nodes $r,s \in N$ where an O-D pair is referred to as the pair of an origin node and destination node, (r,s). Each O-D pair is indexed by $k \in K$ and the total number of O-D pairs is $|K|$. In the fixed demand case, there is an amount of demand wishing to travel between each O-D pair denoted by d_k. In the case of variable demand (elastic demand), the volume of travel demand between O-D pair k is defined as a function of the minimum O-D travel cost, μ_k, through the demand function $d_k = D_k(\mu_k)$

The number of trips from O-D pair k using link a is defined as x_a^k. Thus, the total volume of traffic on link a can be defined as $v_a = \sum_{\forall k} x_a^k$. The cost of travelling on link a is defined by a user cost function $c_a(\mathrm{v})$ where the travel time on link a is $t_a(\mathrm{v})$. Similarly, the link flows, v_a, can also be defined as an aggregation of the path flows. Let $p \in P$ be a path and F_p the path flow. Define $\delta_{p,a}$ as a dummy variable where $\delta_{p,a} = 1$ if link a is used by path p and 0 otherwise. Thus, the aggregated link flow can be defined as $v_a = \sum_{\forall p} \delta_{p,a} \cdot F_p$. This flow aggregation can also be written in the vector form. Also, under the assumption of additive path cost the path cost can also be defined as $C_p = \sum_{\forall a} \delta_{p,a} \cdot c_a(\mathrm{v})$.

Note that the user cost may be a combination of travel time and other costs (e.g. tolls). The travel cost is defined as a generalised travel cost (in time units) and the conversion of other costs into the travel time unit is done through the concept of value of time.

In the network, travel demand passing through the network must satisfy the flow conservation constraint. There are two different forms for defining the flow conservation based on either path flows or multicommodity link flows. For the path flow based formulation, let $p \in P$ be a path and $\Delta_{p,k} = 1$ if path p connects the O-D pair k and 0 otherwise. The origin to destination demands can be satisfied with the constraint $d_k = \sum_{\forall p} \Delta_{p,k} \cdot F_p \quad \forall k$ for the fixed demand case and $D_k(\mu_k) = \sum_{\forall p} \Delta_{p,k} \cdot F_p \quad \forall k$ for the elastic demand case.

For the multicommodity link flow based formulation, let $b_i = 0$ if node i is neither the origin nor destination node, and equal to $-d_k$ and d_k if node i is the origin and destination node of O-D pair k respectively. Each node has $|K|$ flow conservation constraints as follows:

$$\sum_{\forall j|(j,i)\in A} x_{(j,i)}^{k} - \sum_{\forall j|(i,j)\in A} x_{(i,j)}^{k} = b_i \quad \forall i \in N; \forall k \in K.$$

For the elastic demand case, $b_i = 0$ if node i is neither the origin nor destination node, and equal to $-D_k(C_k)$ and $D_k(C_k)$ if node i is the origin and destination node of O-D pair k respectively. The summary of the notation used is provided in Appendix C. Next, the definition and different form of users' equilibrium is presented. Throughout the next section, the path-based formulation will be employed. However, it should be noted that the same equilibrium condition can also be framed in the form of multicommodity link flows.

4.2.2 Different formulations of UE condition

Definition of Wardrop's user equilibrium condition

Before presenting different formulations of the user's equilibrium condition, this section defines the standard mathematical form of the user's equilibrium condition. The travellers in the network are assumed to be rational in the sense that they wish to choose their routes so as to minimise their generalised travel costs. The decision taken for different travellers influences the outcome of other travellers' choices through the congestion effect in the network. Wardrop (1952) defines the principle of user's equilibrium condition (UE) characterising the possible outcome of the travellers' route choice decisions in the network. For each O-D pair, let $\mu_k = \min_{\forall p \in P_k} C_p$ be the minimum travel cost from all paths connecting O-D pair k where P_k is the set of paths relevant to O-D pair k. The UE condition for the fixed demand case is such that:

$$\begin{aligned} &F_{p\in P_k} \geq 0 \Rightarrow C_p = \mu_k && \forall p; \forall k \\ &F_{p\in P_k} = 0 \Rightarrow C_p \geq \mu_k && \forall p; \forall k \\ &d_k = \sum_{p\in P_k} F_p && \forall k \end{aligned} \qquad \text{Equation 4-1}$$

This condition states that path p connecting between an O-D pair k will be used by the travellers if and only if the cost of travelling on this path is the minimum travel cost between that O-D pair. This condition is consistent with the definition of the UE condition in which if the path with non-minimum travel cost is used, the travellers on that path will still have an incentive to shift to other paths offering a lower travel cost. Similarly, the UE condition can also be defined for the elastic demand case with an additional condition on the equilibrium between demand and supply following the normal economic equilibrium:

$$\begin{aligned}
&F_{p\in P_k} \geq 0 \Rightarrow C_p = \mu_k && \forall p;\forall k\\
&F_{p\in P_k} = 0 \Rightarrow C_p \geq \mu_k && \forall p;\forall k\\
&d_k \geq 0 \Rightarrow \mu_k = D_k^{-1}(d_k) && \forall p;\forall k\\
&d_k = 0 \Rightarrow \mu_k \geq D_k^{-1}(d_k) && \forall p;\forall k\\
&d_k = \sum_{p\in P_k} F_p && \forall k
\end{aligned} \qquad \text{Equation 4-2}$$

In brief, the first and second conditions in Equation 4-2 represent the condition that the travellers will only choose the cheapest routes for their journey. The second and third conditions imply that the traveller will travel between an O-D pair *k* if and only if the benefit of the trip (represented by the value of the inverse demand function) is at least equal to the travel cost between that O-D pair.

The concept of UE relies on the key assumption of the travellers have perfect knowledge in which all travellers are assumed to be perfectly aware of all possible routes and have accurate information about the travel costs under equilibrium. This assumption may sound too strong. Various alternatives to this UE concept have been proposed to weaken this assumption, e.g. the concept of Stochastic User Equilibrium (SUE).

Despite its strong assumption, the UE condition has played a central role in transport modelling and planning. The concept has been used to develop many transport modelling software tools such as SATURN (Van Vliet, 1982) which has been widely used as a tool to aid transport planning in the real world. The basic definition of the UE above, both for the fixed and variable demand case, can be reformulated into different mathematical forms which will be discussed next.

Background on the formulation of an equilibrium condition

Before defining specifically the forms of UE condition, various useful definitions, properties, and theorem for constructing the mathematical forms of the UE condition are reviewed. Firstly, the concept of a variational inequality which is central to the study of equilibrium modelling is reviewed.

Definition 4-1 Let Ω be a nonempty and closed subset of $\Re^m$ and let F be a continuous function from Ω to itself. The finite dimensional variational inequality problem, denoted by *VI*(Ω,*F*), is to find a vector $y^* \in \Omega$ such that:

$$F\left(y^*\right)^T \cdot \left(y - y^*\right) \geq 0 \quad \forall y \in \Omega$$

In most cases, Ω is a closed convex set (or polyhedron). In geometric terms, the *VI* states that the vector y^* will be the solution to the *VI*(Ω,*F*) if and only if $F\left(y^*\right)$ is orthogonal to the feasible set Ω at the point y^*.

Definition 4-2 The function F from Ω to itself is said to be strictly monotone over Ω if and only if:

$$\left(F(x) - F(y)\right)^T \cdot (x - y) > 0 \ \forall x, y \in \Omega \text{ and } x \neq y.$$

Theorem 4-1 [Nagurney (1993)] Suppose that *F(y)* is strictly monotone on Ω. Then the solution to *VI*(Ω,*F*) is unique.

The concept of a *VI* is closely linked to the concept of a complementarity problem and a nonlinear optimisation problem which is discussed in turn below.

Definition 4-3 Let Ω be the nonnegative orthant in $\Re^m$ and let F be a continuous function from $\Re^m$ to $\Re^m$. The complementarity problem (CP), denoted by *CP*(Ω,*F*), is to find a vector $y^* \in \Omega$ such that:

$$F\left(y^*\right) \geq 0 \text{ and } F\left(y^*\right)^T \cdot y^* = 0.$$

Proposition 4-1 Let Ω be the nonnegative orthant in $\Re^m$. Then $y^* \in \Omega$ solves the problem *VI*(Ω,*F*) if and only if it also solves the problem *CP*(Ω,*F*).

The proposition above shows the *CP*(Ω,*F*) is a special case of *VI*(Ω,*F*), but note that the converse is not generally true. Notice that definition of Ω for *CP*(Ω,*F*) which is basically the nonnegative orthant in $\Re^m$. This is too restricted assumption. However, the formulation and interrelationship between *CP* and *VI* can also be extended to some specific case. In particular, if the Ω is a polyhedron which is defined by a set of linear equality and inequality constraints, then *VI*(Ω,*F*) can be converted into a mixed complementarity problem (*MCP*).

Proposition 4-2 [Tobin (1986)] Let Ω be a polyhedron defined as

$\Omega = \left\{ y \in \Re^m : Ay = b, Cy \geq d \right\}$ where A is a matrix with the dimension of $l \times m$ and $b \in \Re^l$; C is a matrix of dimension $n \times m$ and $d \in \Re^n$. Then y^* solves the *VI*(Ω,*F*) if and only if for some $x^* \in \Re^n, z^* \in \Re^l, \left(y^*, x^*, z^*\right)$ solves the following *MCP*:

$$\begin{matrix} free \\ 0 \leq \\ free \end{matrix} \begin{pmatrix} y \\ x \\ z \end{pmatrix} \perp \begin{pmatrix} F(y) - C^T x - A^T z \\ Cy - d \\ Ay - b \end{pmatrix} \begin{matrix} = 0 \\ \geq 0 \\ = 0 \end{matrix}$$

, where $a \perp b$ means $a^T \cdot b = 0$ or vector a is orthogonal to vector b. Obviously, with the assumption on strong monotonicity of F the MCP above have the unique solution following Proposition 4-1 and Theorem 4-1 described earlier.

As shown, there is a strong connection between the *VI*(Ω,*F*) and *MCP*. The other key reformulation of the *VI* is its equivalent nonlinear optimisation problem under the assumption on the function *F*.

Theorem 4-2 If *F(y)* is a continuously differentiable function on Ω which is a closed and convex set and its Jacobian matrix

$$,\nabla F(y)=\begin{bmatrix} \frac{\partial F_1}{\partial y_1} & \cdots & \frac{\partial F_1}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial y_1} & \cdots & \frac{\partial F_m}{\partial y_m} \end{bmatrix},$$

is symmetric and positive semidefinite, then y^* is the solution to the $VI(\Omega,F)$ if and only if it is the solution to the following nonlinear optimisation problem:

$$\min_y f(y)$$
$$s.t.\ \ y \in \Omega$$

, where $\nabla f(y)=F(y)$.

Equivalent formulations of the UE condition

With the definitions of *VI*, *MCP*, and its equivalent optimisation problem (*EO*) discussed earlier we are ready to state the UE condition (for elastic demand condition) as a *MCP*, *VI*, and *EO* as follows.

The UE condition stated in Equation 4-2 implies that $F_p\ \forall p \in P$ satisfies Equation 4-2 if it satisfies the following *MCP* condition:

$$\begin{aligned}
&\left(C_p\left(F_p\right)-\mu_k\right)\cdot F_p=0, && \forall p\in P_k;\forall k\in K\\
&C_p\left(F_p\right)-\mu_k\geq 0, && \forall p\in P_k;\forall k\in K\\
&\left(\mu_k-D_k^{-1}\left(d_k\right)\right)\cdot d_k=0, && \forall p\in P_k;\forall k\in K\\
&\mu_k-D_k^{-1}\left(d_k\right)\geq 0, && \forall p\in P_k;\forall k\in K\\
&\sum_{p\in P_k}F_p=d_k, && \forall k\in K\\
&\mathbf{F}\geq 0,\mathbf{D}\geq 0,\mathbf{d}\geq 0
\end{aligned} \qquad \text{Equation 4-3}$$

Theorem 4-3 A pair of vectors $\left(\mathbf{F}^*,\mathbf{d}^*\right)\in\Omega$, where Ω is the set of feasible path flows and O-D flows, satisfies the UE condition if and only if it satisfies the variational inequality problem

$$\mathbf{C}\left(\mathbf{F}^*\right)^T\cdot\left(\mathbf{F}-\mathbf{F}^*\right)-\mathbf{D}^{-1}\left(\mathbf{d}^*\right)^T\cdot\left(\mathbf{d}-\mathbf{d}^*\right)\geq 0\ \text{ for } \forall\left(\mathbf{F},\mathbf{d}\right)\in\Omega$$

and similarly we can also write the variational inequality with the link flow vector instead of path flow vector under the assumption that the path cost is an additive cost function of link cost following the notations given earlier:

$$\mathbf{c}(\mathbf{v}^*)^T \cdot (\mathbf{v}-\mathbf{v}^*) - \mathbf{D}^{-1}(\mathbf{d}^*)^T \cdot (\mathbf{d}-\mathbf{d}^*) \geq 0 \text{ for } \forall (\mathbf{v},\mathbf{d}) \in \Omega .$$

Note that the cost function adopted in both the *MCP* and *VI* formulations above does not need to be separable, i.e. the travel cost on link *a* does not only depend on v_a. Furthermore, the path-flow based *VI* does not need the path cost operator (C_p) to be additive. This may be useful for some analysis, e.g. nonlinear value of time problem. If the path cost is an additive function of link cost, then the *VI* formulation with link flow vector can be obtained.

The *VI* formulation above can also be stated for the fixed demand case where the second term in the *VI* related to the demand function can be dropped. As explained briefly in the introduction, Smith (1979) was the first to state the UE condition in the *VI* form whose formulation was recognised later by Dafermos (1980) as the *VI*. Following Theorem 4-1, if the link cost function are strictly monotone under the link flow vector, then the *VI* for link flow has a unique UE solution (the feasible space for link flow are compact).

Interestingly, in the literature of traffic equilibrium analysis the first (and probably the most widely used) mathematical form is in fact the equivalent optimisation (*EO*) form as suggested by Beckmann *et al* (1956). Following Theorem 4-2 above, the only condition we need to ensure equivalence between the UE condition and the *EO* problem is the symmetric and positive semidefinite condition of the link cost vector. One could try to prove this condition with the general link cost function, where the cost on link *a* is a function of link flow vector, but in general it is very difficult to get the symmetrical and positive semidefinite condition with this general link cost vector. A particular form of link cost function that makes these two conditions satisfied is the separable link cost function. A link cost function is called to be separable if the cost on link *a* is only a function of the flow on that link.

Theorem 4-4 A pair of vectors $\left(\mathbf{v}^*,\mathbf{d}^*\right)\in\Omega$, where Ω is the set of feasible link flows and OD flows, satisfies UE condition if and only if it is the solution to the following nonlinear optimisation problem:

$$\begin{aligned}&\min_{(\mathbf{v},\mathbf{d})}\sum_a\int_0^{v_a}c_a(x)\,dx-\sum_k\int_0^{d_k}D_k^{-1}(y)\,dy\\&s.t.\;\;(\mathbf{v},\mathbf{d})\in\Omega\end{aligned}$$

Notice that the Jacobian of the objective function of the *EO* formulation is equivalent to the *F* function of the *VI* formulation for UE stated in Theorem 4-3. It should be noted that the *EO* formulation of the UE is probably the form with the strongest assumption including the separable link cost function and additive path cost function. On the other hand, the *EO* formulation is the corner stone of the development of successful algorithms for solving the traffic equilibrium assignment with a large scale problem in the past three decades.

An alternative formulation of the *MCP* for the UE condition using the multicommodity flow instead of path flow can also be formulated. Let *E* be the node link incidence matrix, with the dimension $\Re^{|N|}\times\Re^{|A|}$ and with entries $e_{i,a}$ = -1 if node *i* is the origin of link *a*, 1 if node *i* is the destination of link *a*, and 0 otherwise. Recall that x_a^k is the link flow from OD pair *k* on link *a*. In this context, x_a^k will be referred to as $x_{i,j}^k$ where *i* and *j* are the start and end nodes of link *a* respectively. Let λ_i^k be the 'node price' associated with node *i* and OD pair *k*. Let Λ^k be the node-OD incidence vector, with the dimension of $\Re^{|N|}$, which is zero in all positions but two where the element *n* = 1 if node *n* is destination of OD pair *k* and -1 if node *n* is the origin node of OD pair *k*. The *MCP* formulation using the multicommodity link flows can be defined as follows:

$$\begin{aligned}&x_{i,j}^k\cdot\left[c_a\left(v_a,\tau_a\right)-\left(\lambda_j^k-\lambda_i^k\right)\right]=0 &&\forall(i,j)\in A;\forall k\in K\\&E\cdot\mathbf{v}^k=\Lambda^k\cdot D_k^{-1}\left(\mu_k\right)&&\forall k\in K\\&\mu_k=\left(\Lambda^k\right)^T\cdot\lambda^k&&\forall k\in K\\&x_{i,j}^k\geq0&&\forall(i,j)\in A;\forall k\in K\\&c_a\left(v_a,\tau_a\right)-\left(\lambda_j^k-\lambda_i^k\right)\geq0&&\forall(i,j)\in A;\forall k\in K\end{aligned}\qquad\text{Equation 4-4}$$

where $\mathbf{V}^k$ denotes the vector of multicommodity link flow for OD pair k and λ^k is the vector of node prices for all node for OD pair k. The first condition is similar to the complementarity condition for the path-flow based formulation where a path will be used if and only if path cost is the minimum travel cost between that OD pair. In this case, link a will contain some flow from commodity k (OD pair k) if and only if the travel cost on this link is equal to the difference between the node prices at the origin and destination nodes of that link. The second condition is the flow conservation in node-link style. At all nodes the sum of flow entering that node from OD pair k must be the same as the exit flow for that OD pair except the origin and destination nodes where the exit and entry flow is equal to the flow determined by the corresponding demand function. The third condition determines the minimum travel cost for each OD pair as the difference between the node prices at the origin and destination nodes of that OD pair.

The *MCP* formulation with multicommodity link flow offers some advantage over the path-flow based formulation. In using this *MCP* with the optimal toll design context, all the paths must be generated and included in the formulation of the *MCP* (with the path-flow based formulation). Generating all possible paths in a network may not be practical. A strategy to avoid this computational burden is to include the paths as necessary, but in the optimal toll design problem the active path set may change with the toll design. With the multicommodity flow formulation *MCP*, the number of constraint for *MCP* can be defined as priori since the number of nodes and links are fixed even when the road pricing scheme design changes. This advantage will be exploited later on in the next chapter in developing a method to solve the optimal toll design problem without using path flow information. However, it should be noted that the number of variables and constraints may increase rapidly with the multicommodity flow formulation (the number of variables = $|A|\cdot|K|+|N|\cdot|K|+|\tau|$).

Similarly, the *EO* formulation of UE can also be defined based on the multicommodity flows instead of path flows. The only change needed is to use the flow conservation constraint for the multicommodity flow instead of the path flow based formulation and define link flows as aggregation of multicommodity link flows.

The formulation of UE in the forms of *VI*, *MCP*, and *EO* is probably the most widely used and most useful for our purpose of solving the optimal toll problem. However, there exist other possible formulations of UE, gap function and fixed point condition. As will be described in the next chapter, some of these formulations have also been explored as the possible tool for solving the MPEC and optimal toll design problem.

Definition 4-4 Let S be the set of solutions to *VI*. A function $\Phi(\mathbf{x})$ from X to $\Re$ is a gap function for *VI* if

(i) $\Phi(\mathbf{x})$ is restricted in sign on X, and

(ii) $\Phi(\mathbf{x}) = 0 \Leftrightarrow \mathbf{x} \in S$

The first condition means that if $\Phi(\mathbf{x})$ has the same sign (either negative or positive) for all $\mathbf{x} \in X$. Hearn *et al* (1984) defined UE in the form of gap function. The *VI* for UE condition can also be stated as:

$$\mathbf{c}(\mathbf{v}^*)^T \cdot (\mathbf{v}^* - \mathbf{v}) - \mathbf{D}^{-1}(\mathbf{d}^*)^T \cdot (\mathbf{d}^* - \mathbf{d}) \leq 0 \quad \text{for } \forall(\mathbf{v},\mathbf{d}) \in \Omega$$

, and the gap function for UE condition can be defined as:

$$\Phi(\mathbf{v},\mathbf{d}) = \max_{(\tilde{\mathbf{v}},\tilde{\mathbf{d}})\in\Omega} \left\{ \mathbf{c}(\mathbf{v})^T \cdot (\mathbf{v} - \tilde{\mathbf{v}}) - \mathbf{D}^{-1}(\mathbf{d})^T \cdot (\mathbf{d} - \tilde{\mathbf{d}}) \right\} \qquad \text{Equation 4-5}$$

,where Ω is a set of feasible link and demand flows satisfying the flow conservation constraints. Thus, $\Phi(\mathbf{v},\mathbf{d}) = 0$ if and only if $(\mathbf{v},\mathbf{d})$ is the feasible UE solution and $\Phi(\mathbf{v},\mathbf{d})$ is always greater than or equal to zero (non-negative) for all $(\mathbf{v},\mathbf{d}) \in \Omega$. Indeed, the function $\Phi(\mathbf{v},\mathbf{d})$ is not differentiable everywhere. Hearn *et al* (1984) define the subdifferential of $\Phi(\mathbf{v},\mathbf{d})$ as:

$$\partial\Phi(\mathbf{v},\mathbf{d}) = conv\left\{ \mathbf{c}(\mathbf{v}) - \mathbf{D}^{-1}(\mathbf{d}) + \mathbf{c}'(\mathbf{v}) \cdot (\mathbf{v} - \hat{\mathbf{v}}) - \mathbf{D}'^{-1}(\mathbf{d}) \cdot (\mathbf{d} - \hat{\mathbf{d}}) \middle| (\hat{\mathbf{v}},\hat{\mathbf{d}}) \in \Xi(\mathbf{v},\mathbf{d}) \right\},$$

Equation 4-6

where $\mathbf{c}'(\mathbf{v})$ and $\mathbf{D}'^{-1}(\mathbf{d})$ are the Jacobian of $\mathbf{c}(\mathbf{v})$ and $\mathbf{D}^{-1}(\mathbf{d})$ at $(\mathbf{v},\mathbf{d})$, and $\Xi(\mathbf{v},\mathbf{d})$ is the set of solutions to the problem:

$$\max_{(\tilde{\mathbf{v}},\tilde{\mathbf{d}})\in\Omega} \left\{ \mathbf{c}(\mathbf{v})^T \cdot (\mathbf{v} - \tilde{\mathbf{v}}) - \mathbf{D}^{-1}(\mathbf{d})^T \cdot (\mathbf{d} - \tilde{\mathbf{d}}) \right\}.$$

The proof of the above argument can be found in the seminal text of nonsmooth optimisation[8] by Clarke (1983). Notice that the solution to the problem in Equation 4-5 is simply a solution to a shortest path problem at a given flow vector (v,d) and under some condition on the link cost and path cost the solution to this problem is unique (strongly monotone link cost and additive path cost). $\Phi(\mathbf{v},\mathbf{d})$ can also be referred to as the Frank-Wolfe subproblem (linear programming problem) which is the element for finding the descent direction in the sense of simplical decomposition algorithm (see Lawphongpanich and Hearn(1984)).

Patriksson (1994), page 79, also defines a gap function similar to the function proposed by Hearn *et al* (1984):

$$\Phi(\mathbf{v},\mathbf{d})=\left\{\mathbf{c}(\mathbf{v})^{T}\cdot\mathbf{v}-\mathbf{D}^{-1}(\mathbf{d})^{T}\cdot\mathbf{d}\right\}-\min_{(\hat{\mathbf{v}},\hat{\mathbf{d}})\in\Omega}\left\{\mathbf{c}(\mathbf{v})^{T}\cdot\hat{\mathbf{v}}-\mathbf{D}^{-1}(\mathbf{d})^{T}\cdot\hat{\mathbf{d}}\right\}.$$

The interpretation of this gap function is the difference in the total travel cost between that of the flow vector (v,d) and that of the shortest paths given the costs at (v,d). A positive value of this gap function coincides with the situation where some travellers can still benefit in reducing his/her travel time from switching his/her paths to a shorter path. Thus, the function can be used to measure how far the current solution (v,d) is from the UE solution. Similar to the gap function proposed by Hearn *et al* (1984), the value of this gap function is always non-negative and bound below at zero value (at equilibrium solution). Again, we can also derive the Jacobian of this gap function using the subdifferential of a nonsmooth function. As will be shown later in the next chapter, gap function has recently gained attention from various researchers in developing a practical optimisation tool for solving the MPEC and optimal toll design problem.

4.3 MPEC FORMULATION OF THE OPTIMAL TOLL PROBLEM

4.3.1 MPEC and BLPP framework

As mentioned in the introduction, if one wants to analyse the optimal design of a road pricing scheme, one must include the possible response of the travellers into the analysis in order to represent the real world situation. The framework of the interaction between the transport planner, who is in charge of road pricing scheme design, and the road users

---

[8] Note that if the solution to Equation 4-5 is singleton (i.e. unique), then Equation 4-6 is differentiable.

can be depicted as shown in Figure 4-1 below.

The objective of the planner (in which he/she wishes to maximise/minimise) is a function of the responses from the road users. Of course, the planner could iteratively design and implement the road pricing scheme (based on the current travel pattern), observe the responses from road users to the tolls imposed (following the UE condition), evaluate the benefit of that particular design, and redesign the scheme to optimise his/her objective function with the updated travel pattern, and then go through the process until some form of convergence is satisfied.

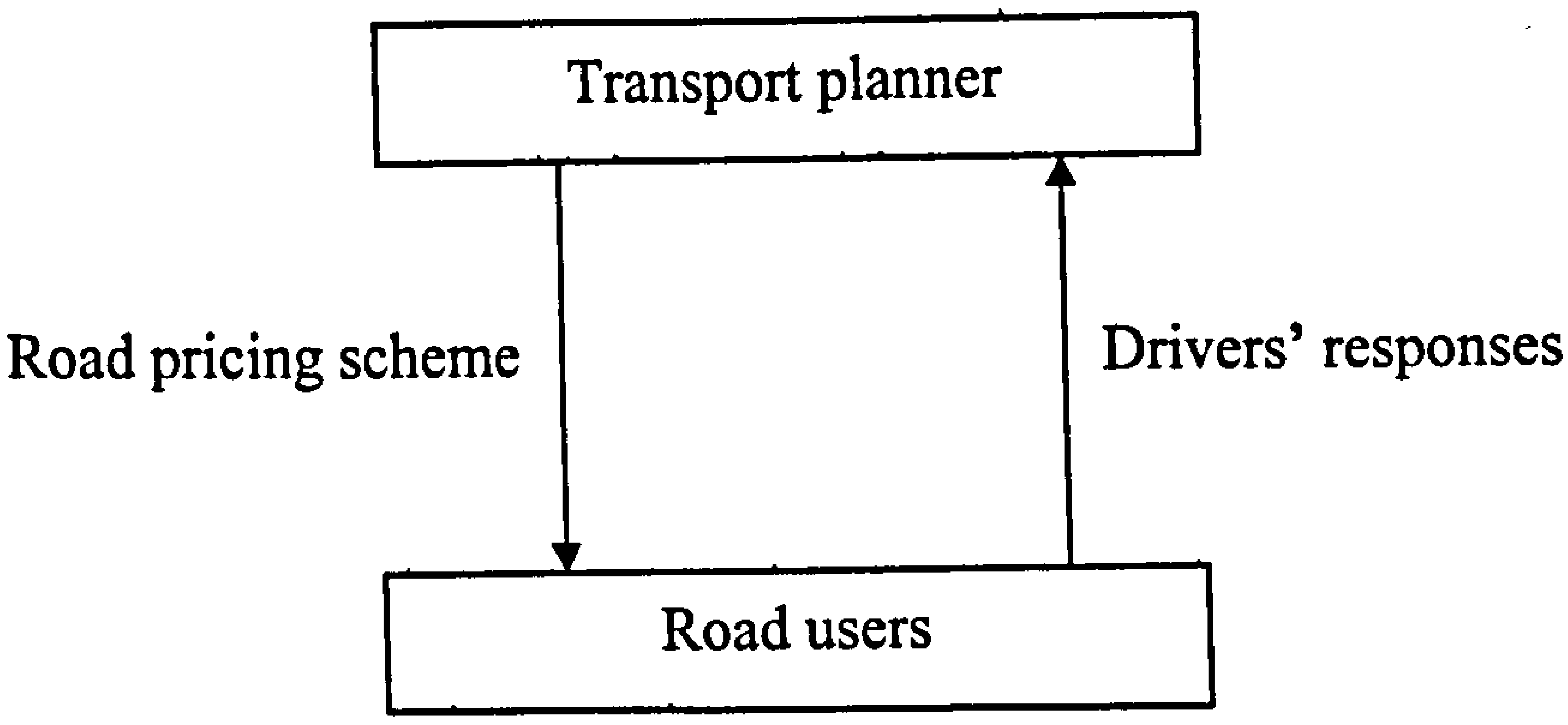


Figure 4-1 Users -Planner interaction

On the other hand, the planner could take into account the potential responses of the road user to the tolls during the design process. In this setting, the planner is assumed to be able to anticipate the responses of drivers to the toll imposed (using the UE condition). With this formulation, the optimal design of a road pricing scheme is the one that is also optimal after allowing road users to respond to the scheme. Note that the first and second approaches mentioned may converge to different solutions. Fisk (1984) discussed the difference between these two interaction processes. The former interaction pattern will converge to a Nash equilibrium condition whereas the latter will converge to a Stackelberg game equilibrium condition.

Mathematically, we can define the Stackelberg game as a Mathematical Program with Equilibrium Constraint (MPEC) (Harker and Pang, 1988; Luo *et al*, 1996). As the name suggests, MPEC is an optimisation problem in which one of the constraints of the problem is an equilibrium condition. Let $x \in \Re^n$ and $y \in \Re^m$ be two sets of variables in

MPEC, f is a function from $\Re^{n+m}$ to $\Re$, and F is a mapping function from $\Re^{n+m}$ to $\Re^{m}$. The function f is the objective function of MPEC and F is the parameterised equilibrium function for the *VI*. Let X be the feasible region of x and $\Omega(x)$ be the feasible set of y given a vector x. The MPEC can be formulated as:

$$\begin{aligned} &\min_{(x,y)} f(x,y) \\ &s.t. \\ &x \in X \\ &y \in S(x), \end{aligned} \qquad \text{Equation 4-7}$$

where $S(x) \equiv \left\{ y^{*} \in \Omega(x) \middle| F(x,y^{*})^{T} \cdot (y - y^{*}) \geq 0, \forall y \in \Omega(x) \right\}$, which implies that for each vector of x *S(x)* is the set of the solutions to the *VI(F,* $\Omega(x)$).

The *VI* representing an equilibrium condition is included as one of the constraints in this optimisation problem explaining the name of MPEC. The equilibrium constraint is usually referred to as the Nash game (Nash, 1951). In a Nash game, there are a number of players (say, M) each of whom can choose his or her strategy in order to maximise (or minimise) his or her economic objective. Each player observes the action of others and then chooses his or her strategy optimally. The equilibrium point is when there is no player who can improve his or her objective by changing the strategy. The *VI* and *MCP* are probably the most popular form of equilibrium condition adopted in MPEC.

The origin of the MPEC is the Stackelberg game (Stackelberg, 1952). The Stackelberg game is different from the Nash game. In the Stackelberg game there is a distinguishable player who can anticipate the reaction of others (the leader) whereas all players in a Nash game are homogeneous. From the perspective of Stackelberg game, f is the payoff function of the leader in the game which is a function of his/her decision variable x and the responses from the followers y. The followers respond to the leader's action following the equilibrium condition cast in the form of *VI*, and as mentioned the leader in this game can anticipate the followers' reactions in which the existence of *VI* as the constraint into the leader's decision problem represents this ability of the leader.

As suggested in Theorem 4-2, under some conditions on the link cost function and path cost function, the *VI* can be formulated as a nonlinear optimisation problem. If the equilibrium condition is defined as an optimisation problem, the MPEC becomes a bilevel optimisation programming problem (BLPP). BLPP can be described as an optimisation program with a constraint region determined by a series of optimisation problems. The BLPP is a special class of multilevel programming which was first defined by Candler and Norton (1977). It is widely accepted that BLPP and MPEC is a NP-Hard[9] problem. Until present there is no exact optimisation method that can guarantee the global optimal solution (Jeroslow, 1985; Ben-Ayed and Blair, 1990). The review of the algorithms for solving MPEC and BLPP will be given in the next chapter.

The structure of MPEC provides a good platform for formulating the optimal road pricing design problem mathematically following the idea in Figure 4-1. Next the optimisation problem for the optimal road pricing design is formulated as an MPEC.

4.3.2 Variations of optimal road pricing design problems

There are various factors and elements concerning the design of a road pricing scheme following the review in Chapter 2 and the survey in Chapter 3. The problem discussed in this thesis is limited to the case of a single user class, single time period, and single mode. Thus, only some of the design elements and factors of a road pricing scheme are actually considered in the problem formulation including:

- determination of the toll levels
- determination of the toll point locations/number of toll points
- constraints on the toll level/pattern (i.e. uniform or variable)
- constraints on the toll location and pattern (i.e. cordon format)
- constraints on the possible adverse impact and expected benefit (outcome constraint)

In the optimal road pricing design problem, the road users are assumed to respond to the road pricing scheme following the UE condition explained earlier. The planner wishes to optimise some objective functions (e.g. social welfare, revenues, environmental improvement).

---

[9] The NP-Hard problem is a very complex decision problem in which it can be solved at least in a polynomial time on a non-deterministic Turing machine.

Let $\psi_i(\tau,\varepsilon,\mathbf{v},\mathbf{d})$ $i=1,\ldots,M$ be the objectives of the planner where τ is the vector of tolls and ε is the binary vector (1-0) representing the toll location in which $\varepsilon_a=1$ if link a is tolled and 0 otherwise. $\mathbf{v}$ and $\mathbf{d}$ are the vectors of link flows and demand flows. Let Θ be the set of feasible vector of ε. The generalised formulation of the optimal road pricing design problem can be stated as:

$$\begin{aligned}
&\max_{(\tau,\varepsilon,\mathbf{v},\mathbf{d})}(\psi_1,\ldots,\psi_M)\\
&s.t.\\
&0\le\tau_a\le\varepsilon_a\cdot\bar{\tau}_a \qquad \text{for } \forall a\in A\\
&\underline{\mathbf{b}}\le\mathbf{H}\cdot\tau\le\bar{\mathbf{b}}\\
&\mathbf{g}(\mathbf{v},\mathbf{d})\le\mathbf{0}\\
&\varepsilon\in\Theta\\
&(\mathbf{v},\mathbf{d})\rightarrow sol\big(VI\big(F(\mathbf{v},\mathbf{d},\tau,\varepsilon),\Omega(\tau,\varepsilon)\big)\big)
\end{aligned}$$

Equation 4-8

where $\mathbf{H}$ is an $\Re^h\times\Re^A$ matrix and represent possible constraints on the toll level and pattern, $\mathbf{g}$ is a vector function from $\Re^{A+K}$ to $\Re^l$ representing possible nonlinear and linear constraint related to adverse impact and/or expected benefit. For brevity, $\mathbf{g}$ will be referred to as the functions for the outcome constraints of the design. An example for the application of H is, say:

$$\begin{bmatrix}0\\0\end{bmatrix}\le\begin{bmatrix}1&-1&0\\0&1&-1\end{bmatrix}\cdot\begin{bmatrix}\tau_1\\\tau_2\\\tau_3\end{bmatrix}\le\begin{bmatrix}0\\0\end{bmatrix}, \text{ implying that } \tau_1=\tau_2=\tau_3.$$

The first constraint implies that link a will be tolled if and only if $\varepsilon_a=1$ where $\bar{\tau}_a$ is the upper bound of the toll level on link a. The constraint $\varepsilon\in\Theta$ implies that the vector of tolled links must be a member of a feasible set which is loosely defined at this stage (e.g. in Chapter 6 Θ is defined more precisely as the set of tolled links forming a charging cordon, or alternatively it can represent the set of possible tolled links and prohibited links for tolling, or the possible number of tolled links). Finally, the last constraint represents the condition that the pair of vector $(\mathbf{v},\mathbf{d})$ must be the solution of the $VI\big(F(\mathbf{v},\mathbf{d},\tau,\varepsilon),\Omega(\tau,\varepsilon)\big)$ given the road pricing design vector (τ,ε) representing the UE condition.

Some specific forms of the optimal road pricing design problem can also be stated based on this generalised problem formulation including the continuous optimal toll design problem (given the toll location) and the mixed-integer optimal toll problem (decide on toll location, number of toll points, and toll level). In fact, it is straightforward to define different optimal road pricing design problems, but to ensure the clarity of the discussion in the following chapters the problems are stated clearly in this section.

Continuous optimal toll problem (COTP)

The continuous optimal toll design problem (COTP) is probably the simplest problem from the family of optimal road pricing design problems defined in Equation 4-8. Nevertheless, it is still a very complex optimisation problem given the nature of the MPEC formulation. The problem involves selecting an optimal toll level for each pre-defined tolled link in the network. The problem can be stated as:

$$\begin{aligned} &\max_{(\tau,\mathbf{v},\mathbf{d})} \psi_1(\tau,\mathbf{v},\mathbf{d}) \\ &s.t. \\ &0 \le \tau_a \le \varepsilon_a \cdot \bar{\tau}_a \qquad \text{for given } \varepsilon \text{ and } \forall a \in A \\ &(\mathbf{v},\mathbf{d}) \rightarrow sol\left(VI\left(F(\mathbf{v},\mathbf{d},\tau),\Omega(\tau)\right)\right) \end{aligned} \qquad \text{Equation 4-9}$$

The problem stated above can also be associated with multiobjective and additional constraints on the toll levels and outcome constraints. In addition, any form of UE discussed earlier in 4.2 can be used as the constraint of the optimal toll design problem (to replace *VI*). Different forms of UE may lead to different solution algorithms which will be discussed fully in the next chapter.

Mixed 1-0 optimal toll problem (MOTP)

The other variation of the optimal road pricing design problem is the mixed 1-0 optimal toll design problem (MOTP). The design variables in the MOTP include the locations of the toll points, the number of toll points, and the toll level for chosen tolled links. The inclusion of the 1-0 variable (ε) representing the state if link a is tolled or not makes the problem become a mixed 1-0 optimisation problem. The problem can be stated as follows:

$$\begin{aligned}
&\max_{(\tau,\varepsilon,\mathbf{v},\mathbf{d})} \psi_1(\tau,\varepsilon,\mathbf{v},\mathbf{d})\\
&s.t.\\
&0 \le \tau_a \le \varepsilon_a \cdot \bar{\tau}_a \quad \text{for } \forall a \in A\\
&(\mathbf{v},\mathbf{d}) \rightarrow sol\left(VI\left(F(\mathbf{v},\mathbf{d},\tau,\varepsilon),\Omega(\tau,\varepsilon)\right)\right)
\end{aligned} \qquad \text{Equation 4-10}$$

Indeed, the problem is very complex given the MPEC and the combinatorial nature of the problem. The location problem of the optimal road pricing design is a rarely investigated research topic. Despite the importance and relevance of the topic to practice, surprisingly only few researchers have tried to develop an algorithm to tackle the problem (see Verhoef, 2002; Shepherd and Sumalee, 2004). The objective functions and the outcome constraints of the optimal road pricing design are only stated in arbitrary form in Equation 4-8. The next section identifies and formulates some possible indicators that could be adopted as the objectives or outcome constraints.

4.4 POSSIBLE OUTCOME INDICATORS

The original concept of road pricing was to enhance the economic efficiency of the transport system. In other words, road pricing was proposed as a fiscal tool for market intervention so that the social welfare was maximised. However, as reviewed in Chapter 2 and from the survey results in Chapter 3, road pricing was found to serve a wider range of objectives. The main criterion for the development of the algorithm for designing a practical/optimal road pricing scheme is to accommodate a wide range of objectives individually or simultaneously (these issues are addressed later in Chapter 7 and Chapter 8 where the cordon designs are associated with constraints and multiple objectives respectively).

As will be shown later on, the optimisation algorithm developed could accommodate arbitrary forms of objective functions and constraints. The aim is simply to state clearly in this section the objective functions/constraints and their mathematical forms that will be used later on in the thesis for the numerical tests. The objectives/constraints selected are the most widely used indicators in designing or evaluating a road pricing scheme based on the survey results in Chapter 3.

It should be clarified that objectives can be treated as constraints and vice versa. From Equation 4-8, we can notice that there exist a number of objectives and outcome constraints. For instance, the design outcome of the net revenue generated can either be considered as an objective function or as an outcome constraint during the design process. Thus, in this section we refer to both objectives and constraints as indicators.

Five main indicators are adopted throughout the thesis including:

- Net economic benefit
- Spatial equity
- Net revenue
- Total travel time
- Total travel distance

The mathematical formulations for the last three indicators are relatively simple and will not be discussed here. Instead, we concentrate on the formulation of the first two indicators.

Social welfare improvement

For each OD pair, the response of demand to the travel cost (including the toll) is captured in the form of the demand function, D_k. The demand function represents the aggregated willingness to pay of the road users for accessing the destinations (i.e. this implies the benefit of the users for reaching those destinations). In general, the lower the cost the higher the travel demand (trips). From the demand function and a given number of trips made, User Benefit (UB) can be defined as the area under the inverse demand curve (which is a function of the number of trips made):

$$UB = \sum_{\forall k} \int_0^{d_k} D_k^{-1}(x)dx$$

As explained in Section 4.2.1, we assume that the congestion in the network happens solely on the links and is a function of the traffic volume on the link. Given a volume of travel demand, the total Social Cost (SC) incurred from the trips is:

$$SC = \sum_{\forall a} v_a \cdot t_a(v_a)$$

It should be stressed that the cost involved in the expression of *SC* excludes the cost from the toll imposed. The social welfare function or economic benefit (*EB*) is then defined as the net difference between the *UB* and *SC*:

$$EB = UB - SC = \sum_{\forall k} \int_0^{d_k} D_k^{-1}(x)dx - \sum_{\forall a} v_a \cdot t_a(v_a).$$

This can be illustrated as shown in Figure 4-2. In Figure 4-2, the equilibrium point between the demand and supply curves is at C trips and E costs.

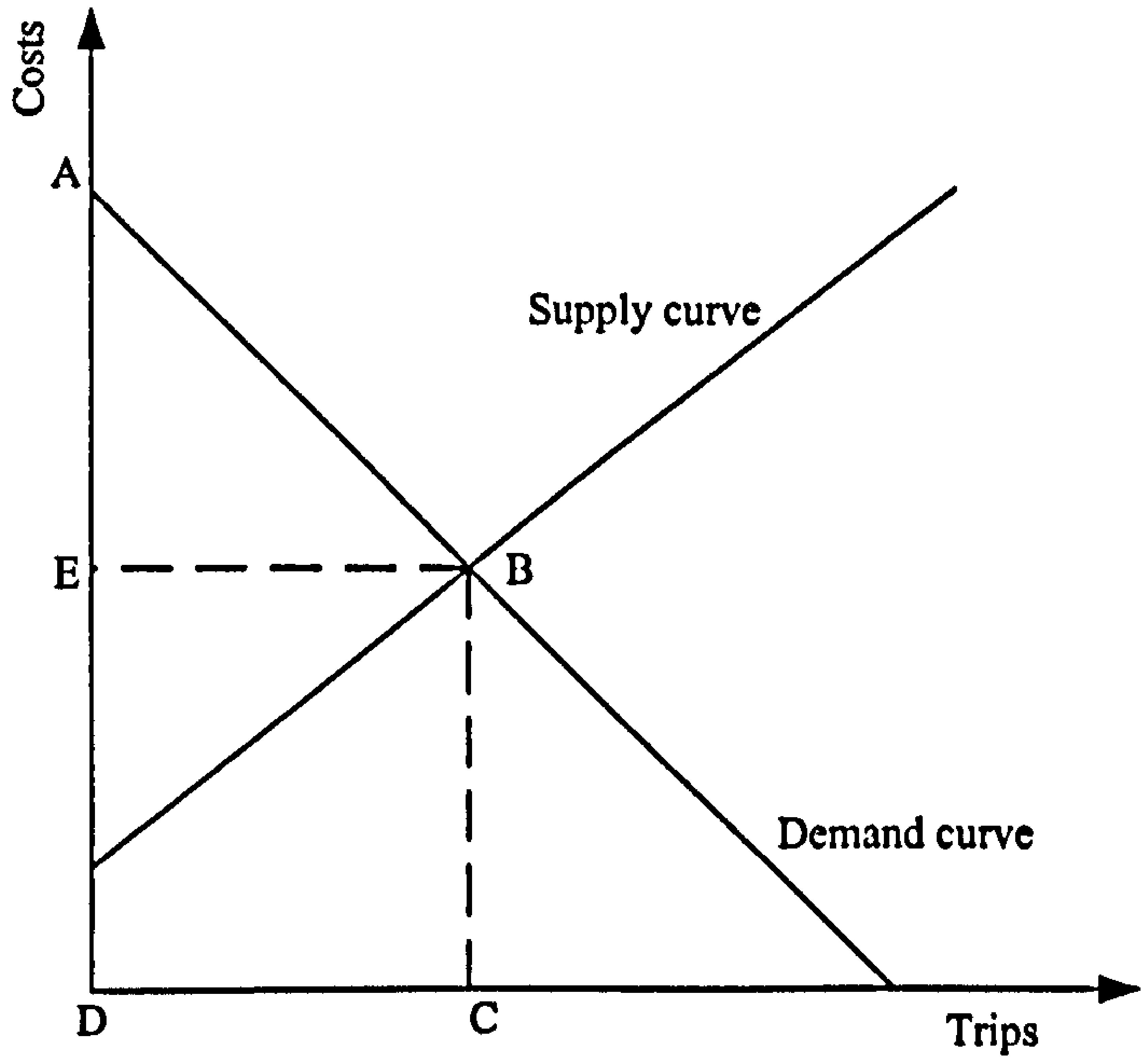


Figure 4-2 Graphical illustration of the definition of economic benefit

The *EB* is, as explained, the area under the demand curve from zero trip to the trip at the equilibrium point (area A-B-C-D-A). The cost incurred from the trips (*SC*) is represented by the area E-B-C-D-E. The economic benefit is therefore the area A-B-E-A.

The other cost associated with the road pricing scheme is the operator costs, *OC* (the costs imposed on the toll operator, e.g. local authority). Let α_a be the costs for implementing and operating a toll point on link *a* (per the same unit of time as defined for *UB* and *SC*). The net economic benefit (net EB) can then be defined as:

$$\text{net } EB = EB - OC = \sum_{\forall k} \int_0^{d_k} D_k^{-1}(x)dx - \sum_{\forall a} v_a \cdot t_a(v_a) - \sum_{\forall a} \varepsilon_a \cdot \alpha_a \qquad \text{Equation 4-11}$$

Note that in most cases, the value of α_a is the same for all links in the network. The other way to evaluate the net *EB* is to concentrate on the improvement of the net *EB* compared to the no-toll scenario. Let d_k^0 and v_a^0 be the demand flow between OD pair *k* and link flow on link *a* in the no-toll scenario. The *EB* improvement can be defined as:

$$\left[\sum_{\forall k}\int_0^{d_k} D_k^{-1}(x)dx-\sum_{\forall a}v_a\cdot t_a(v_a)-\sum_{\forall a}\varepsilon_a\cdot\alpha_a\right] -\left[\sum_{\forall k}\int_0^{d_k^0} D_k^{-1}(x)dx-\sum_{\forall a}v_a^0\cdot t_a\left(v_a^0\right)\right] \qquad \text{Equation 4-12}$$

Figure 4-3 below exemplifies the meaning of the formulation of the net *EB* improvement. In this figure, there are two supply curves (one with toll and one with no toll). The number of trips at the equilibrium points for the no toll and toll cases are C and G respectively (associated with travel cost of E and H excluding the toll).

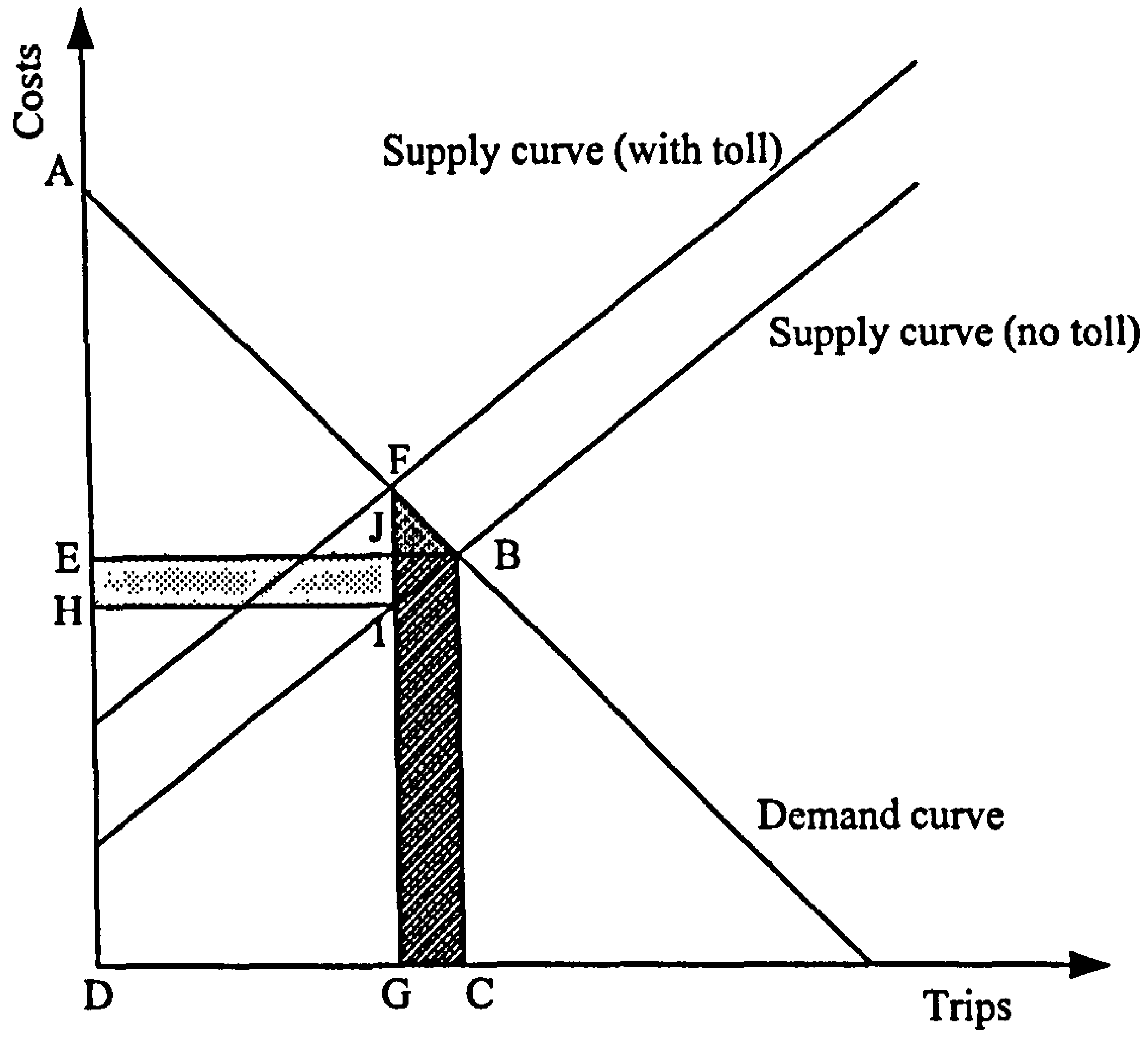


Figure 4-3 Graphical illustration of the definition of economic benefit improvement

The area F-B-C-G-F (shaded area) is the difference between the *UB* of the no toll and toll cases (negative value). The change in SC is the area E-B-C-G-I-H-E (dotted area) which is positive value. Thus, the *EB* improvement is the difference of the area E-J-I-H-E (positive value) and the triangle F-B-J-F (negative value). Of course, the net *EB* improvement is the *EB* improvement minus the cost of the road pricing scheme.

The other form of the *EB* improvement is based on the rule of half for calculating the change in consumer surplus (ΔCS):

$$\text{net } EB=\Delta CS+revenue-OC=\Delta CS+\sum_a\tau_a\cdot\varepsilon_a\cdot v_a-\sum_a\alpha_a\cdot\varepsilon_a \qquad \text{Equation 4-13}$$

, where $\Delta CS = \frac{1}{2}\cdot\left\{\sum_{k}\left[\left(d_k^0 + d_k\right)\cdot\left(\mu_k^0 - \mu_k\right)\right]\right\}$, d_k^0 and d_k are the OD demand in the no toll and toll cases respectively, and μ_k^0 and μ_k are the OD travel cost (including tolls) in the no toll and toll cases in that order. Figure 4-4 illustrates the interpretation of this formula and proves the equivalence between the two different forms of net *EB* improvement.

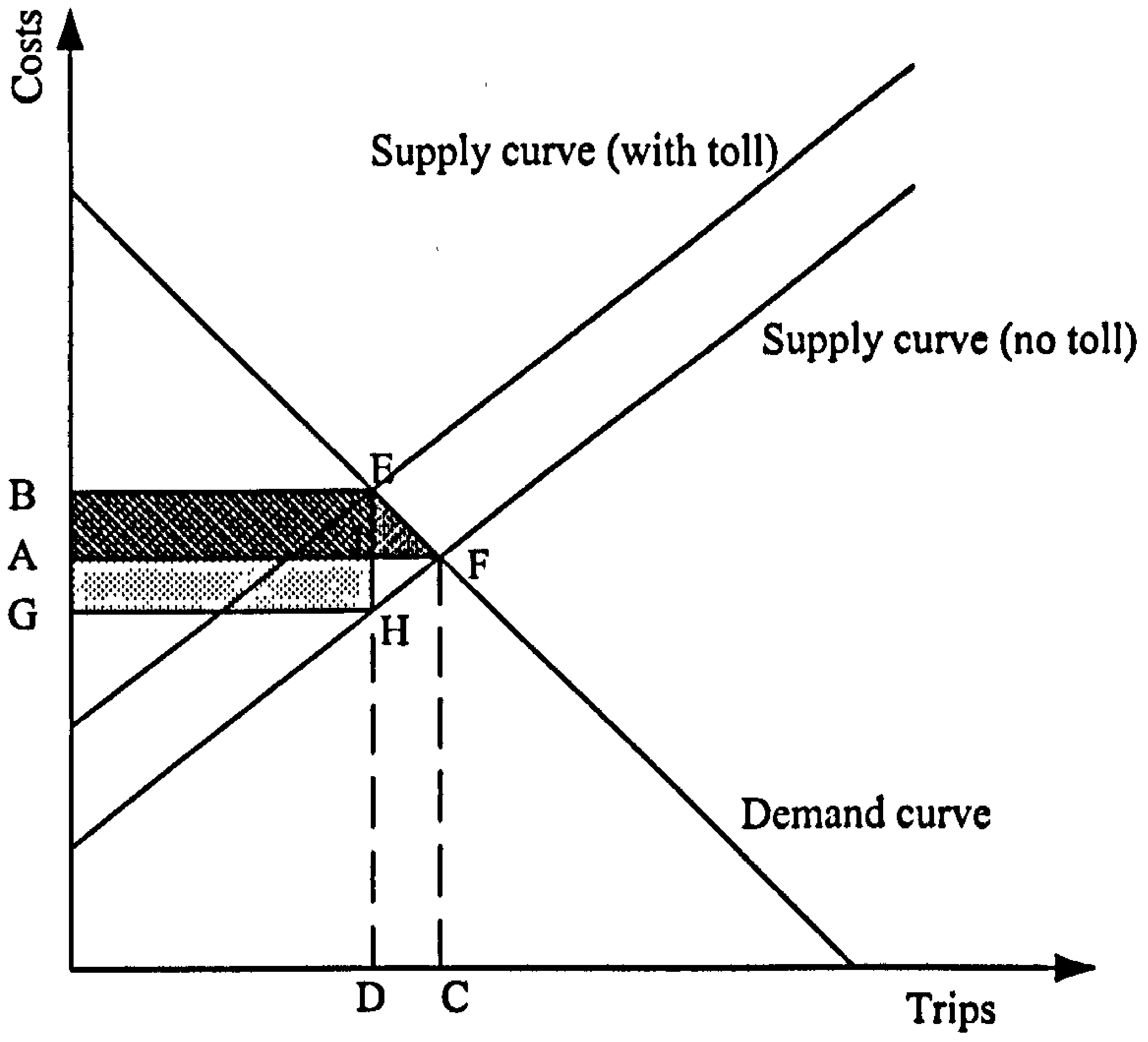


Figure 4-4 Graphical illustration of the definition of economic benefit improvement based on rule of half

From Figure 4-4, the toll level implemented is B-G, ΔCS is the area of the trapezium B-E-F-A-B (based on rule of half and is a negative value), and the revenue generated is the square B-E-H-G-B. Thus, the *EB* improvement is the difference between these two shapes that are the area A-I-H-G-A (positive value) and the triangle E-F-I-E (negative value). This is exactly the same of the result in Figure 4-3. Again, the net *EB* improvement is the *EB* improvement minus the OC. In the case with a very nonlinear shape of the demand function and/or a significant level of the tolls, the net EB improvement as calculated by the rule of half may underestimate the net *EB* improvement (mainly due to of the area E-F-I-E not being triangular).

One possible drawback of using the absolute net EB is the restriction of the demand function. An example of the demand function that may cause a problem in calculating the UB is:

$$d_k = d_k^0 \cdot \left(\frac{\mu_k}{\mu_k^0} \right)^{\beta},$$

where β is the own elasticity (which should be negative) and its inverse demand function is:

$$\mu_k = \mu_k^0 \cdot \left(\frac{d_k}{d_k^0} \right)^{\frac{1}{\beta}}.$$

With this inverse demand function if one tries to integrate from zero trip to trip d_k, one would get an improper integral. However, if the integral is from non zero trip to d_k, the above demand function is applicable to the economic evaluation (the case of net *EB* improvement).

There is a long established consensus on the form of the social welfare improvement or the net economic benefit improvement as shown in this section. Several studies on the design of optimal road pricing have adopted the objective function in Equation 4-11 (Yang and Huang, 1998; Verhoef, 2002) and will be used in this thesis.

Spatial equity

The most criticised drawback of road pricing is its potential equity impact. There are two main dimensions of the equity impact: vertical and horizontal equity impact. This is also supported by the evidence in Chapter 3 where a number of local authorities express their concern over the equity impact of the road pricing scheme. The vertical equity impact is concerned with the unequal impact of road pricing on different groups of the population (e.g. classified by income level, gender, access to car, etc.). The horizontal equity impact is also referred to as the '*spatial equity impact*'. The spatial equity impact can be described as the distribution of the benefits and costs of the scheme across the population from different areas in the network. If the scheme benefits only a small group of people from a few areas, but the rest of the population experience a decline in the social welfare, the scheme can be argued as an inequitable policy.

In this thesis, we only consider *'the horizontal dimension'* of the equity impact, given the modelling assumption made earlier (single user class). Some researchers have attempted to incorporate the equity consideration (spatial equity) into the area of network optimisation (where optimal toll is a special case of this problem). Meng and Yang (2002) recently proposed a framework to considering the spatial distribution impact of a change in the traffic network structure (i.e. network design problem). They proposed a measure which is the ratio of the origin-destination (O-D) travel cost in the "*before*" and "*after*" scenarios. Also, Yang and Zhang (2002) use a similar indicator to study the toll design problem with multiple user classes. Unfortunately, these measures cannot capture the loss in consumer surplus from the depressed demand and the unequal distribution of the cost/benefit of the scheme.

There exists an economic framework that can be utilised in analysing the spatial equity effect which is an indicator for income inequality. The *Gini coefficient* is the most commonly used income inequality measure. In fact, Fridström *et al* (2000) used the same measure to analyse the distribution effect of various transport policies. It can be explained with reference to Figure 4-5 below.

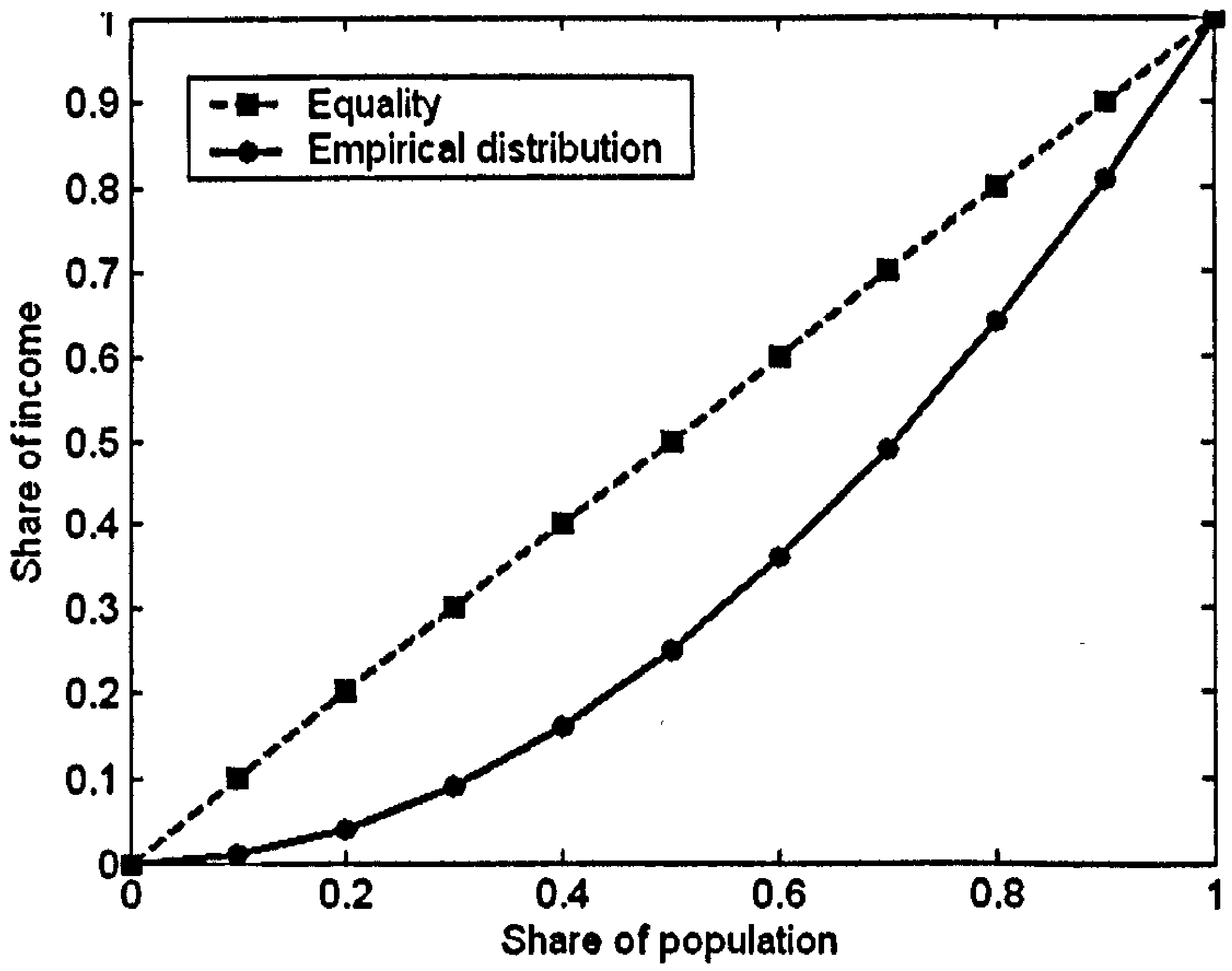


Figure 4-5 Lorentz curve explaining the distributional impact

On the horizontal axis, a population is ordered by income from the lowest to the highest. On the vertical axis is the cumulative share of total income. The ideal situation or the most equitable case is that everybody has the same income and this will produce the straight line. In reality, the cumulative income distribution will not form a straight line. This is shown by the '*Empirical distribution*' curve or so called '*Lorentz curve*'. The area between the two curves represents the skewness of the actual income distribution from the ideal equitable distribution. Thus, this area can be used as an indicator of income inequality, ranging from 0 for perfectly equal income distributions to 0.5 for distributions where one person earns all income (which is the area of the triangle under the equitable income distribution line). The *Gini* coefficient is twice this area to get a measure of inequality between 0 and 1.

In our case, the main component of our interest is the spatial distribution of costs and benefits of the road pricing scheme. Although there are different ways to measure the distribution of the impact (e.g. absolute change or relative change), the absolute change in social welfare for each O-D pair *k* is used as the measure for the distribution impact. The main drawback of the relative change is that it will increase the inequality of the benefit distribution in the absolute term, i.e. the group with already high benefit requires a higher level of benefit to achieve the same level of relative improvement compared to the group with low benefit in the do-nothing situation. The disaggregated social welfare improvement (without operator cost) for each OD pair will be used as the measure of the distribution of benefits and impacts for the traveller group from that O-D pair *k*:

$$\hat{\psi}_k(\mathbf{F},\mathbf{d},\tau) = \psi_k(\mathbf{F},\mathbf{d},\tau) - \psi_k^0(\mathbf{F}^0,\mathbf{d}^0) = \left(\int_0^{d_k} D_k(x)dx - \mu_k \cdot d_k\right) - \left(\int_0^{d_k^0} D_k(x)dx - \mu_k^0 \cdot d_k^0\right) \qquad \text{Equation 4-14}$$

where **F** is a vector of path flows with the entries F_p, and **d** is a vector of OD demand with the entries d_k (the superscript 0 denote the no-toll scenario). Note that the main difference between this formulation and the formulation presented earlier is that this formulation uses the OD flows as variables. The reason for doing this is to enable the calculation of the disaggregated social welfare improvement for each OD pair.

The other key assumption is the revenue recycling. As discussed, in the formulation of the net *EB* improvement the social cost does not include the user cost incurred by the tolls (i.e. only the travel time is used to calculate the user costs). This implies that the revenue is recycled fully back to the users which can be interpreted as shown in

Equation 4-13 that when the tolls are included as the user cost in ΔCS, the revenue is already added back to the calculation.

For the analysis of the distributional impact of the road pricing scheme, it is considered to be difficult to make any assumption on revenue recycling. One could assume that the users from each OD pair receive the revenue recycled back according to what they pay. This is probably a very difficult assumption to justify. The other alternative is to actually decide on the approach to distribute the revenues to the users (e.g. construction of the new mass transit line), but this involves another model and simulation to analyse the proportion of benefit the users from different OD pairs actually receive. This discussion over an appropriate assumption on this is definitely important but beyond the scope of this thesis. In this thesis, a rather simple assumption is adopted where the revenues are not recycled back to the users at all. Thus, the only benefit and cost incurred to the users from each OD pairs is the loss in user benefit and the travel time saving. With this definition, the actual analysis of the distribution impact turns out to concentrate on the value of ΔCS. Thus, $\hat{\psi}_k(\mathbf{F},\mathbf{d},\tau)$ in Equation 4-14 becomes the ΔCS for OD k.

For our purposes, probably the most useful formulation of the *Gini* coefficient is:

$$\psi_{equity} = \frac{1}{2\cdot\left(d^0\right)^2\cdot\bar{\psi}_{welfare}}\cdot\sum_{l=1}^{K}\sum_{k=1}^{K} d_l^0\cdot d_k^0\,|\hat{\psi}^l_{welfare}-\hat{\psi}^k_{welfare}| \qquad \text{Equation 4-15}$$

, where $\bar{\psi}_{welfare}$ denotes the average of ΔCS from all OD pairs. There are totally K O-D pairs with ΔCS $\psi_{welfare} = \left(\hat{\psi}^1_{welfare},\ldots,\hat{\psi}^K_{welfare}\right)$, d_k^0 is the total number of travel demand (i.e. number of trips) of group k in the un-tolled case, $k = 1, \ldots, K$ and $\sum_{\forall k} d_k^0 = d^0$ is the total travel demand in the whole network.

However, the ΔCS can be both negative and positive. This contrasts to the sign of the income (used in the calculation of the original *Gini* coefficient) which can only be a positive value. Thus, the raw value of the ΔCS for each O-D pair will need to be re-scaled before calculating the *Gini* coefficient. Let $\hat{\psi}^J_{welfare}$ be the lowest value of the ΔCS

(from O-D pair *j*) from all values (which may be negative). Then, the ΔCS for the other O-D pairs can be re-scaled as follows:

$$\widehat{\psi}^k_{welfare} = \hat{\psi}^k_{welfare} + \left|\hat{\psi}^j_{welfare}\right| + \kappa \qquad \forall k$$

, where κ is a given positive constant. Also, the average ΔCS will be replaced by the average of the re-scaled ΔCS, $\widehat{\psi}^k_{welfare}$. It is easy to verify that this re-scaling will not change the scale of the *Gini* coefficient since the main purpose is to measure the area under the equitable curve. As mentioned, the value of the *Gini* coefficient is between 0-1. The lower the *Gini* coefficient, the more equitable the road pricing scheme.

Of course, other interpretations and formulations of the spatial distribution impact of the road pricing scheme also exist and could be more appropriate than the one adopted here. The main focus of the thesis is rather to demonstrate the ability of the optimisation algorithm developed later on in Chapter 6 to Chapter 8 to deal with a less conventional form of outcome indicator. Obviously, the *Gini* coefficient in Equation 4-15 already poses several mathematical problems, e.g. it is a non-smooth function and may not be differentiable at some point. Thus, the optimisation algorithm developed must be able to deal with this kind of mathematical formulation.

4.5 SUMMARY

This chapter provides necessary background information on the optimal road pricing design problem discussed later on in the thesis. The framework of the optimal road pricing design problem is formulated as a Mathematical Program with Equilibrium Constraints (MPEC) which can also be considered as a bilevel optimisation programming problem (BLPP) or Stackelberg game. The lower level or equilibrium constraint is the Wardrop's user equilibrium (UE) condition representing the responses of travellers to the road pricing scheme. The model adopted in this thesis is limited to the case of a single user class, single mode, and single time period. Different formulations of UE are discussed and presented in the chapter including the variational inequality (VI), mixed complementarity problem (MCP), equivalent optimisation (EO), and gap function.

From the review and survey results presented earlier in Chapter 2 and Chapter 3, several key components that are necessary to be included in the optimal road pricing design were identified (including the toll location, toll level, toll pattern, form of toll location, outcome constraint, and multiple objectives). A generalised optimisation problem for the optimal road pricing design taking into account these features is presented. In particular, two specific problems of the road pricing scheme design are discussed, the continuous optimal toll problem, and the mixed 1-0 optimal toll problem. The first problem involves the decision on the optimal toll level given a set of tolled links. The second problem includes the toll location and number of toll points as the decision variables as well as the optimal toll level. These two problems will be the main objects for the discussion in the next chapter. The treatment of other elements in the optimal road pricing design (e.g. format of toll point, multiple objectives, and outcome constraint) will be discussed later on in Chapter 6 to Chapter 8.

Two main outcome indicators, social welfare and equity, which can be used either as an objective function or as a constraint are discussed and formulated mathematically. The social welfare is defined as the net economic benefit improvement over the no-toll case. For equity, the modelling approach adopted in this thesis only allows the consideration of the spatial distribution impact (over different areas of the network). The index of *Gini* coefficient commonly employed for studying income distribution is modified to represent the distribution of impact of the cost/benefit of road pricing scheme on the users from different OD movements.

The following four chapters will tackle these aspects of optimal road pricing scheme design. The next chapter will concentrate on solving a more theoretical setting problem (optimal toll level or second best toll problem). Then, Chapter 6 integrates the constraint on the topology of a road pricing scheme into the design process. Chapters 7 and 8 then consider the outcome constraint and multiobjective problems respectively.

CHAPTER 5 OPTIMISATION ALGORITHM FOR OPTIMAL TOLL PROBLEM

Nothing at all takes place in the universe in which some rule of maximum or minimum does not appear. Leonhard Euler

5.1 INTRODUCTION

In Chapter 4 the continuous optimal toll design problem (COTP) was formulated as a Mathematical Program with Equilibrium Constraints (MPEC):

$$
\begin{aligned}
&\min_{(\tau,\mathbf{v},\mathbf{d})} \psi_1(\tau,\mathbf{v},\mathbf{d}) \\
&s.t. \\
&0 \le \tau_a \le \varepsilon_a \cdot \overline{\tau}_a \qquad \text{for given } \varepsilon \text{ and } \forall a \in A \\
&(\mathbf{v},\mathbf{d}) \rightarrow sol\left(VI\left(F(\mathbf{v},\mathbf{d},\tau),\Omega(\tau)\right)\right)
\end{aligned}
$$

where the equilibrium condition in the problem is the Wardrop's user equilibrium condition.

In this chapter, an extensive review of the algorithms for tackling the COTP is presented. Firstly, the development of algorithms from the transport research area is discussed. The algorithm for solving COTP is closely linked with the algorithm for the so called network design problem (NDP). In the NDP, the design variables are link capacity, new links, toll level, and toll point. Thus, COTP can be seen as an instance of the NDP.

Apart from the review of developments in the transport area, this chapter will also present a taxonomy[10] of the possible solution algorithms for the MPEC from the optimisation research area which is applicable to the COTP.

After the review, three possible optimisation methods for solving the COTP are presented and tested with different networks. The first method is based on the idea of a

---

[10] Note that the reader who is more interested in the algorithms proposed in this chapter can skip the reviews in Sections 5.3 and 5.4.

merit function which is used to smooth the complementarity condition of the UE, referred to as the smoothing algorithm. The complementarity condition representing the UE, as discussed in Chapter 4, poses two problems with the nonlinear optimisation problem. The first is its non-differentiability and the second is the linear-dependent condition at all feasible points. These two problems are the main obstacles to the development of the efficient optimisation algorithm for MPEC. The complementarity condition of the UE is smoothed by a particular form of the merit function, the perturbed Fischer-Burmeister function. This smoothed version of the complementarity constraint eliminates these two problems mentioned earlier and allows the application of any off-shelf nonlinear optimisation algorithm to solve the problem. The smoothing algorithm is implemented with a Sequential Quadratic Programming (SQP) algorithm and tested with a five-link network.

The second method is the improved version of the cutting plane algorithm (CPA) originally proposed by Lawphongpanich and Hearn (2004). CPA treats the UE as a variational inequality (*VI*) and exploits the polyhedral structure of the feasible flows so that the feasible region can be presented by only a set of extreme points of the polyhedron. In the original CPA, the polyhedron structure is only exploited in the lower level formulation (UE condition). In the improved CPA presented in this chapter, the polyhedron structure of the feasible region of link/demand flows is exploited further in the upper level. The advantage of this implementation is a great reduction in the number of constraints and variables involved in the COTP. The improved CPA is implemented and tested with a small network.

The last method is the Genetic Algorithms (GA) based method. GA is a meta-heuristic optimisation algorithm which does not rely on the mathematical properties of the functions involved in the optimisation problem. GA-CHARGE is developed to tackle the COTP.

The structure of the chapter is as follows. Section 5.2 reviews the development of solution algorithm to the MPEC problem in the transport area. Section 5.3, then, provides a wider perspective of which solution algorithms for MPEC have been developed in the optimisation area. Obviously there is some overlap between these two sections. Section 5.4 presents a smoothing algorithm for solving the COTP and the test results with a small network (five link network). Section 5.5 presents the improved CPA algorithm and test results from a small network (Pacman network). Section 5.6 proposes GA based algorithms for solving COTP (GA-CHARGE) and test results from the Edinburgh network. Finally, Section 5.7 concludes the chapter.

5.2 DEVELOPMENT OF ALGORITHMS IN THE TRANSPORT AREA

The optimal toll design problem is an instance of the network design problem (NDP). Thus, the reference to COTP and NDP are exchangeable in this review. The problem of NDP was first proposed by Steenbrink (1974). Since then, there have been many key developments in this area, see Figure 5-1. Note that the summary in Figure 5-1 is by no means a comprehensive list of the research in the NDP area. The figure instead aims to summarise the key developments during the past three decades.

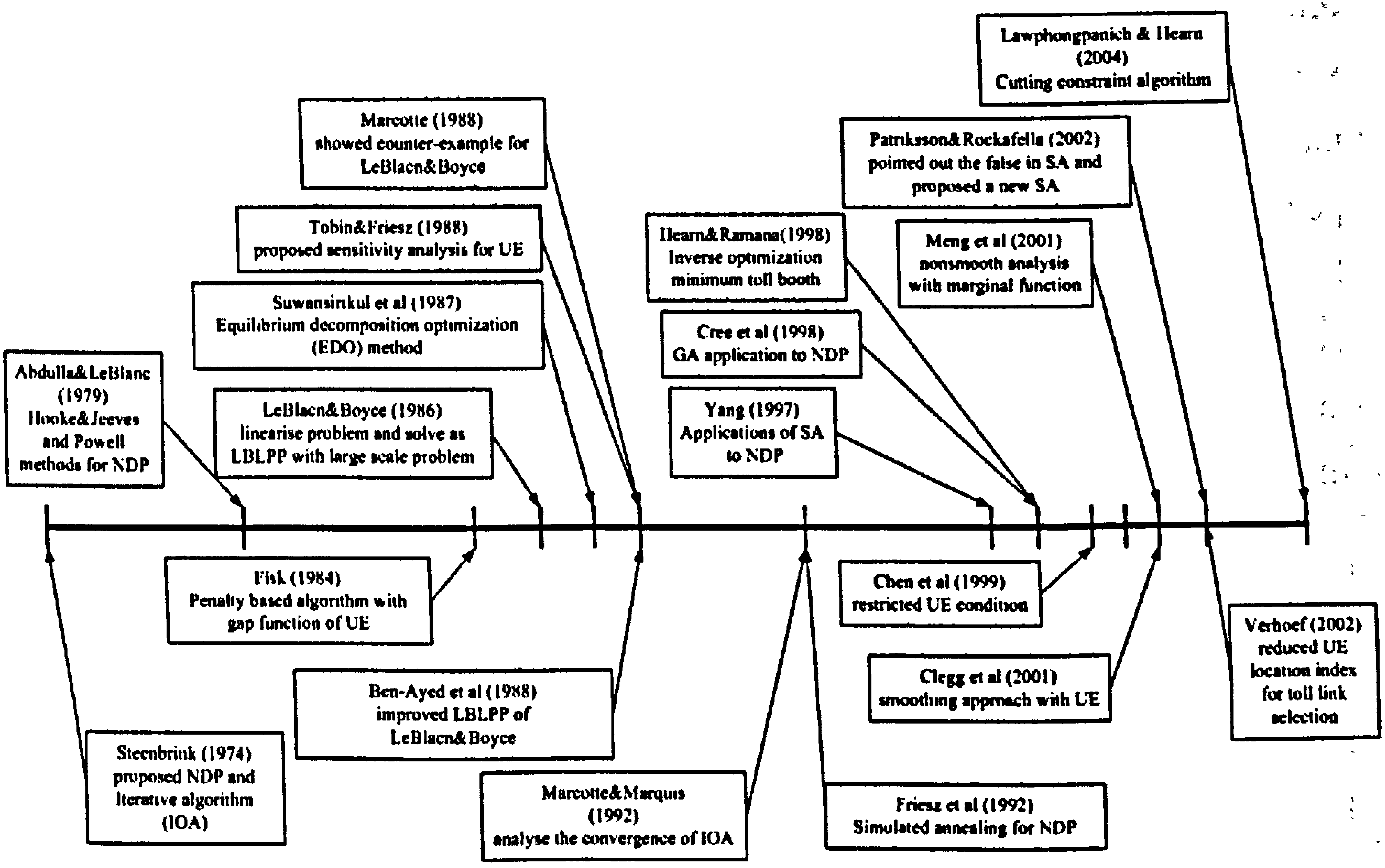


Figure 5-1 Milestone of the development of the transport NDP

As shown in Figure 5-1, the area of NDP and MPEC problems in transport has received consistent attention from the transport research community since the 1970's. Throughout the timeline, various approaches from different traits of mathematical and operations research have been proposed and applied to the NDP. Good evidence of the complexity of this problem is probably the number of imperfect algorithms proposed over the years that have generated counter-examples later on.

The different algorithms proposed for solving the transport NDP can be categorised into eight main groups according to their strategies for dealing with the UE constraint (see Chapter 4 for a discussion of the different formulations of UE):

(i) Iterative optimisation approach (Steenbrink, 1974; Abdulaal and LeBlanc, 1979; Suwansirikul *et al*, 1987)

(ii) Linearization approach (LeBlanc and Boyce, 1986; Ben-Ayed *et al*, 1988; LeBlanc and Boyce, 1988)

(iii) Sensitivity analysis (Tobin, 1986; Yang, 1997; Patriksson and Rockafellar, 2002)

(iv) Inverse optimisation (Hearn and Ramana, 1998)

(v) Minmax/maxmin based optimisation (Fisk, 1984; Meng *et al*, 2001)

(vi) Smoothing approach and reduced UE condition (Chen *et al*, 1999; Clegg *et al*, 2001; Verhoef, 2002)

(vii) Cutting plane algorithm (Lawphongpanich and Hearn, 2004)

(viii) Meta-heuristic optimisation approach (Friesz *et al*, 1992; Cree *et al*, 1998)

The categories suggested above represent the current state-of-the-art of optimisation algorithms for solving the transport NDP. Next, each category of the approaches for solving the transport NDP is discussed.

5.2.1 Iterative optimisation approach

The first approach proposed for solving the NDP is the iterative optimisation method where the method alternates between solving the NDP without UE condition (referred to as the upper level) and the UE problems (referred to as the lower level). In other words, the method first solves the upper level with fixed solution of the lower level and then subsequently solves the lower level with fixed solution of upper level. The process will repeat until it reaches convergence (e.g. the change of the link flows between the two iterations is lower than the tolerance value).

Steenbrink (1974) proposed the Iterative Optimisation Assignment (IOA) that involves the process between solving the upper problem with fixed travel response and solving the lower level with fixed design alteration. However, Marcotte and Marquis (1992) analysed the convergence of the IOA method and illustrated that the convergence of this method is not necessarily to a local optimum. In addition, Fisk (1984) and Friesz and Harker (1985) suggested that IOA, if it converges, leads to the Nash's equilibrium point between the leader and followers rather than the Stackelberg equilibrium.

Suwansirikul *et al* (1987) applied an equilibrium decomposition optimisation (EDO) heuristic with the NDP. This method is operated similarly to the IOA but with the inclusion of the Stackelberg leader-follower structure. For the EDO, the NDP is decomposed into interacting subproblems. Each subproblem is the travel cost on a link where the problems interact since the flows on each link depends on the vector of link improvements of all links. The main drawback is that the methods cannot guarantee the convergence to the global or even local optimum. In a rather different manner, Abdulaal and LeBlanc (1979) applied two heuristic derivative free optimisation methods, Powell and Hook&Jeeves methods, to the NDP and also discrete NDP. Again, the convergence of these algorithms to a local optimum is not guaranteed.

5.2.2 Linearisation approach

The second strategy proposed for solving the NDP is to linearise the NDP problem resulting in a linear bilevel optimisation programming problem (LBLPP) and then solve it by using a global algorithm for solving the LBLPP. LeBlanc and Boyce (1986) developed the BLPP for the network design problem in the fixed demand case with user-optimum driver behaviour. They applied a piecewise linear approximation to the total travel cost function on each link. Thus, the optimisation program for user equilibrium becomes a LBLPP and they employed a simplex-based algorithm for the LBLPP as proposed by Bard (1983). However, their formulation does not allow the inclusion of a concave form of the investment function. This assumption is unrealistic in representing economies of scale in link improvements.

Ben-Ayed *et al* (1988) later formulated the LBLPP for the network design problem under fixed demand based on the linearization of both the objective function in the upper level and of the user equilibrium program (lower level). Their paper improved the formulation proposed by LeBlanc and Boyce (1986) further by allowing both concave and convex link improvement functions and also allowed a greater generality in representing the travel cost functions. An interesting comment in their paper is that the method used to solve LBLPP based on Bard (1983) is not always able to solve the optimisation problem. Marcotte (1988) later confirmed the inability of this method to solve LBLPP. He used a simple two-link counterexample to show that the solution from this method is not the solution for LBLPP.

The drawback of the method is the accuracy of the approximation of the total cost on each link. The method relies heavily on the piecewise linear approximation which requires the appropriate number of segmentation of the total travel cost function. This is not a trivial task and can lead to inaccurate results. Furthermore, the accuracy of the method to solve LBLPP is uncertain. Nevertheless, there have been a number of newly proposed global methods for solving the LBLPP efficiently (see Section 5.3.1). Thus, there exist some possibilities for reviving the linearisation approach as a possible way for tackling the NDP.

5.2.3 Sensitivity analysis

The third strategy for tackling the NDP is to derive the sensitivity expression of the user equilibrium condition to the change of the upper level decision variables. Then, this sensitivity expression (a linear function) can be used to represent the users' responses to the network design variable instead of the UE condition; hence reducing the problem to a normal optimisation problem. This method was proposed by Tobin and Friesz (1988) following the seminal work of Fiacco (1976). Friesz *et al* (1990) applied sensitivity analysis to the development of heuristic solution methods for the network design problem. Yang (1997) derived the sensitivity expression for the user equilibrium condition under elastic demand. He also applied the method to the problem of network design and optimal toll problem. Yang and Bell (1997) looked into the problem proposed by Ferrari (1995): to seek a link toll pattern on a road network so as to hold traffic demand within a given level such as a network environmental, or physical, capacity. They also employed the sensitivity expression to change the NDP to a single level optimisation program.

The weakness of this method is that the range of the design variables has to be limited in order to ensure the accuracy of the approximation of the response surface. The sensitivity expression also depends on the differentiability of the equilibrium condition. This condition is very much subject to the assumption of the strict complementarity constraint of the KKT condition of the UE (which may not be true). Tobin and Friesz (1988) proposed a method to avoid the problem of the non-unique path flows by using the extreme point of the set of feasible path flows. However, recently Patriksson and Rockafellar (2002) suggested that the differentiability of the equilibrium condition is not guaranteed. They also proposed a different method for deriving sensitivity of UE flows with respect to the design variable based on the generalised derivative of the UE flows respect to the design variable (directional derivative). The method was applied to a number of NDP with small networks.

5.2.4 Inverse optimisation

Bergendorff *et al* (1997) and Hearn and Ramana (1998) showed that given a vector of system optimal (SO) link flows, the toll vectors that can make the system optimal link flows to be UE flows can be defined as a polyhedron implying that system optimal link tolls are not unique. Given the set of feasible link tolls, an optimisation program can be formed to select the best link tolls to produce SO link flows, e.g. minimise the revenue (minimum revenue problem) or minimise the number of toll points. This kind of approach fits in with the concept of inverse optimisation. From the original idea of the inverse optimisation, a solution to an optimisation problem is defined (**x**), then an inverse optimisation problem can be formed to find the solution to this inverse problem such that the **x** is still the optimal solution to the original problem.

Hearn and Yildirim (2002) later extended the approach of Hearn and Ramana (1998) to the optimal toll problem with elastic demand. The method associated with existing standard optimisation software is proved to solve the problem of a modest sized urban network successfully, but the implementation with a larger network still needs more development. Hearn *et al* (2001) exploited the theorem produced in Hearn and Yildirim (2002) to develop a cutting plane algorithm for solving a large scale problem of the minimum revenue pricing framework.

The drawback of the approach used by Bergendorff *et al* (1997); Hearn and Ramana (1998); and Hearn and Yildirim (2002) is the dependency on the user optimal flow interpreted as the objective to maximise social welfare (for the elastic demand case) or to minimise the total travel cost (for the fixed demand case). This mean the design must always consider the objective of social welfare or total travel cost as the main objective and the others as the secondary objectives. This restriction reduces the flexibility of changing the objective functions for road pricing schemes. In addition, the method is not suitable for the problem with pre-defined tolled links (and low number of tolled points), since the solution to the inverse optimisation problem may not exist.

5.2.5 Minmax/maxmin based optimisation

As explained in the previous chapter, the UE condition can be included into the NDP as *VI*, *MCP*, *EO*, or gap function. The methods in this category employed different forms of gap functions to represent the UE condition in the NDP. The gap function is a function representing the distance of the present solution from the equilibrium solution. The gap function will only equal to zero at an equilibrium solution.

Fisk (1984) proposed a penalty based method with the gap function as shown in Equation 4-5 in the last chapter. The difficulty in using the gap function is its form as a maximisation or minimisation problem. The concept of subdifferential can be used to define the derivative of the min or max function following Dem'yanov (1990) and, as explained in the previous chapter, if the solution to the maximisation or minimisation problem of the gap function is unique then the gap function is differentiable. For the problem in Equation 4-5, if the strong monotone condition on the link cost function and demand function is assumed, then following Theorem 4-1 the problem in Equation 4-5 has a unique solution. Thus, the penalty based function proposed in Fisk (1984) under the strong monotone condition of link cost and demand function became a differentiable optimisation problem despite a non-standard form of constraint.

Surprisingly, despite a nice property of the gap function based approach there has not been any other development of the optimisation algorithm for the NDP for a practical sized network. Until recently Meng *et al* (2001) proposed a continuously differentiable problem with a new form of gap function. In their paper, they refer to the reformulated gap function as the "marginal function". The marginal function adopted is based on the *EO*, see Theorem 4-4 in the previous chapter, of the UE condition. Let $\varphi(\beta,\mathbf{v},\mathbf{d})$ be the objective function of the *EO* of UE, the marginal function can be defined as:

$$\xi(\beta,\mathbf{v},\mathbf{d})=\varphi(\beta,\mathbf{v},\mathbf{d})-\min_{(\mathbf{v}^*,\mathbf{d}^*)\in\Omega}\varphi(\beta,\mathbf{v}^*,\mathbf{d}^*).$$

Thus, this marginal function has all the properties of a gap function. Meng *et al* (2001) showed the Jacobian of the marginal function above which is similar to the Jacobian of the gap function in Equation 4-6 and implemented it with an augmented Lagrangian algorithm. The algorithm was tested and compared with other optimisation approaches (including EDO, Simulated Annealing, and Hook&Jeeves algorithm). Similar to the gap function, the marginal function is differentiable with the strongly monotone link cost and demand function. In addition, the condition of separable link cost function is also needed for the *EO*. An important advantage of the marginal function based approach which is worth mentioning is that it operates on the link flow level. This helps reduce the number of variables in the problem significantly and provides a good potential for a large scale problem.

5.2.6 Smoothing approach and reduced UE condition

The UE condition can also be expressed as *MCP* (see Section 4.2.2). However, the complementarity condition of the UE is not differentiable everywhere and poses an unwanted condition for the optimisation problem (linear dependence of the active constraints). The methods in this category attempt to use MCP as the UE condition in the NDP but with different treatments on the MCP to avoid the non-differentiability and linear-dependent problem.

Verhoef (2002) analysed the property of the second-best toll pricing problem. He defines the UE condition as *MCP* with path flows, see Equation 4-3. The key assumption was imposed that only the used paths are included in the optimisation problem. Thus, the *MCP* was reduced to a normal nonlinear equality constraint:

$$\begin{array}{ll} \left(C_p\left(F_p\right)-\mu_k\right)=0, & \forall p\in P_k;\forall k\in K \\ F_p>0 & \forall p\in P_k;\forall k\in K \end{array}.$$

Similarly, Chen *et al* (1999) used a similar strategy to reduce the UE condition to a number of equality constraints. They assumed, again, that only used paths are included in the problem and the set of used paths will not change. Then, the UE condition can be defined such that all paths between the same O-D pair must have the same cost (implies the restricted UE condition).

Both methods imposed a strong assumption on the restricted path set to reduce the *MCP* to a normal equality constraint. Although both methods were tested successfully with a medium scale network, the converged solution may not be the true optimum and the extension to the full UE condition may not be possible.

An alternative method to rectify the condition of the *MCP* is to smooth out the complementarity condition. Clegg *et al* (2001) defined a smooth function of the UE complementarity function as the square of the complementarity condition with path flows. With this formulation, all used and unused paths could be included into the optimisation problem. This eliminated the drawback of the two methods discussed earlier. They also proposed a descent search direction method where the search direction is a combined vector of the Jacobian of the upper level objective function and the smoothed complementarity condition with some specified weighting scheme. The combination of both Jacobians in the descent search direction is similar to most of the penalty based approach for solving the nonlinear optimisation problem. However, it is well known that the path flow solution to UE may not be unique and this causes some problem with the descent search direction. The regularised function which is the squared Euclidean distance between two path flow vectors was adopted in the method to avoid this problem.

The methods in this category do not rely on the assumption of the strongly monotone function of the link cost and demand function and the link cost can be non-separable. In fact, the method proposed by Clegg *et al (2001)* was tested with the signal optimisation problem which involved non-separable link cost function. Nevertheless, the methods need a pre-defined set of paths and this may be impractical for a large scale problem.

As mentioned, there exist other possible forms of the smoothed complementarity function. In Section 5.4, an algorithm based on a type of smooth function of the complementarity condition is presented and tested with a small network.

5.2.7 Cutting plane algorithm

The other form of the UE is *VI*, see Theorem 4-3 in the previous chapter. The drawback of the *VI* formulation is that the solution to the *VI* must be compared against all possible flows in the feasible region (Ω). As explained previously, Hearn *et al (2001)* proposed a cutting plane algorithm for solving the minimum revenue toll pricing problem. The cutting plane is adopted to iteratively generate the extreme point of Ω as needed. The sub-problem, for generating an extreme point of Ω with given link costs, is a linear program or more specifically a shortest path problem. Ω is a polyhedron (due to the linearly constraint set) and hence all feasible flows in this polyhedron can be represented as a convex combination of the set of extreme points of the polyhedron.

Lawphongpanich and Hearn (2004) recently proposed a cutting plane algorithm for solving the second-best toll problem. The method is based on a similar idea to Hearn *et al* (2001) where a number of extreme points of the Ω will be generated iteratively. The method was tested with a number of large scale problems and showed some encouraging results.

Given the encouraging results of the algorithm, in Section 5.5, the cutting plane algorithm will be implemented and tested with a small network for experimental purpose. The detail of this algorithm will be discussed in Section 5.5 as well.

5.2.8 Meta-heuristic optimisation approach

The last category of the methods proposed for solving the transport NDP is the so-called meta-heuristic optimisation approach. The meta-heuristic optimisation approach does not rely on any mathematical structure of the function involved in the NDP. The main mechanism driving the optimisation process is a simple search operator inspired by different natural based phenomena. For instance, the approach of Simulated Annealing was developed based on the idea of the metallic annealing process (Kirkpatrick *et al*, 1983) and the Genetics Algorithm (GA) was inspired by the evolution process in nature (De Jong, 1975). Indeed, the flexible structure of the meta-heuristic methods allows a direct application to the MPEC and transport NDP. Friesz *et al* (1992) applied a simulated annealing approach to solve the network design problem with the user equilibrium constraint. Xiong and Schneider (1992) applied a cumulative genetic algorithm and a neural network to a transportation network design problem. Cree *et al* (1998) similarly applied a genetic algorithm to the continuous network design problem. Yin (2000) developed the genetic algorithms based approach to solve various transport NDPs and compared its performance with the sensitivity analysis based approach. However, the approach was only tested with a very small network.

The main advantage of this class of method is that it does not require the derivative or restrict the functional form of the objective functions. Both methods are motivated by the natural optimisation process which involves the evolution and the stochastic step during searching. Therefore, the method can avoid converging to the local optimum and seeking the global optimum. However, the cost of these two methods is the time and computer resources required. Further more, the methods in this class cannot guarantee the optimal solution. There exist other meta-heuristic optimisation approaches which can be applied to the NDP including Tabu-search (Glover, 1989), Local-search, Ant-colony, Neural-network, Differential evolution, etc. The flexibility and global search behaviour of the meta-heuristic optimisation approach becomes attractive for the mixed 1-0 optimisation problem. GA is used later on in Section 5.6 in this chapter to solve the COTP.

5.3 A TAXONOMY OF MPEC ALGORITHMS: OPTIMISATION PERSPECTIVE

The MPEC has gained interest not only from the transport area but also from other disciplines. The review in the previous section showed a strong connection between the development of solution algorithms from the optimisation research community and the progress of solution algorithms for the NLP in a transport context. For example, the idea of the linearisation approach for solving the NDP was based on the algorithm for finding the global optimal solution of the LBLPP proposed by Bard (1983). Similarly, the algorithm proposed in Meng *et al* (2001) was influenced by the work of Shimizu *et al* (1997) which was also proposed by Marcotte and Zhu (1997).

Thus, it is worthwhile to summarise the key developments and current trends of the solution algorithm for the MPEC from the optimisation perspective. For similar reviews on the same subject, see Vicente and Calamai (1994) and Dempe (2003). In addition, a number of books on this subject are available including Luo *et al* (1996), Shimizu *et al* (1997), Bard (1998), Outrata *et al* (1998), and Dempe (2002). Interestingly, these textbooks approach the MPEC or BLPP problem from different angles. For instance, Luo *et al* (1996) mainly concentrates on the application of penalty based approaches to MPEC with the KKT condition representing the equilibrium condition. Shimizu *et al* (1997), Outrata *et al* (1998), and Dempe (2002) adopt methods based on the nonsmooth optimisation approach. This shows a great diversity of ideas and developments of the solution algorithms to MPEC and proves that solving MPEC is still a very challenging and active research topic.

5.3.1 Algorithms for the linear and quadratic BLPP

Two important classes of the BLPP are the linear bilevel optimisation programming problem (LBLPP) and quadratic bilevel optimisation programming problem (QBLPP). With the specific structure of the LBLPP, it is proved that the global optimal solution to the LBLPP will definitely be at one of the extreme points of the inducible region (IR)[11] of the LBLPP (see Theorem 5.2.2 page 200 Bard, 1998). Bialas and Karwan (1982) based on this theorem developed the "Kth-best" algorithm for solving the LBLPP. The algorithm is based on the extreme point ranking procedure. Alternatively, the LBLPP can be converted to a single level mixed-linear optimisation problem via the KKT condition of the lower level problem or the equilibrium condition (the combinatorial part occurs at the complementarity condition of the KKT condition). Then, a global optimisation method for solving the mixed-linear optimisation problem can be applied to the LBLPP. Also recently, there are various global optimisation algorithms proposed for the LBLPP such as Branch&Bound, cutting plane and outer approximation (see for example Hansen *et al*, 1992; Tuy *et al*, 1993; White and Anandalingam, 1993).

For the QBLPP, the problem is actually equivalent to a Mathematical Program with Linear Complementarity Constraints (MPLCC). With the interrelationship between the

---

[11] Inducible region (IR) is the intersection of the feasible set of the upper level and the set of reactions from the lower level problem.

VI, MCP, and the EO explained earlier in the previous chapter, it is easy to see that the QLPP is the MPEC with *VI* having the linear cost mapping (*F*), see Definition 4-1. A special characteristic of the QBLPP is that the IR of the problem is a union set of some faces of the polyhedron (piecewise linear set) which is defined by the upper level constraint without considering complementarity constraint. In addition, the local optimal solution of QBLPP must be at one of the extreme points of the piecewise linear IR, if and only if the objective function is concave (convex) for the minimisation (maximisation) problem.

In this case, any method in a similar fashion to the simplex algorithm for linear programming can be adopted to solve QBLPP, e.g. the "Kth-best" algorithm method proposed by Bialas and Karwan (1982) (moving along the descent extreme direction of the IR). The key optimality condition for the simplex-like algorithm is that at an extreme point of the IR if there is no descent extreme ray (direction), then that extreme point is at least a local optimum, which is the same condition for the simplex algorithm.

Unfortunately, if the main objective function is not concave (convex) for minimisation (maximisation) problem, then some of the local optima may not be at an extreme point of the IR. In addition, the local optimal condition based on the non-existence of the extreme ray at the extreme point is no longer a valid condition. Some of the extreme points may satisfy the optimality condition based on the subsistence of the extreme ray but is not a local optimum for the problem (i.e. there exists some descent direction from that point[12]). This point is referred to as a "Local Star Inducible Region" (LSIR) point (Bard, 1998). This is the key problem for developing an efficient algorithm which exploits the structure of the QBLPP.

Bard (1983) proposed the Branch&Bound approach for solving QBLPP that operates mainly on the complementarity constraint (this method can also be applied to a general BLPP/MPEC). Vicente *et al* (1994) proposed the steepest descent based algorithm for solving QBLPP. Zhang and Liu (2001) proposed the extreme point algorithm to solve QBLPP and MPLCC based on the special characteristic of the QBLPP and the descent

---

[12] Descent 'extreme' direction is the descent direction from one extreme point to another adjacent extreme point where as descent direction can be a descent direction from one solution to any other direction in the neighbourhood of that solution which does not need to be an extreme point.

direction strategy. The algorithm allows the iterative point not to be an extreme point (this is to reflect the aforementioned argument that the local optimum may not be an extreme point).

Instead of using a normal descent search algorithm like Vicente *et al* (1994), the algorithm conducts the search based on the adjacent extreme points and extreme direction of the smallest face of the piecewise linear IR containing that iterative point. The search can move from one iteration to another either through the descent extreme point, descent extreme ray, or descent direction (which is derived efficiently from all of the extreme points of the smallest face of the IR containing the iterative solution). Again, the key feature of this method is the efficient way to exploit the linear piecewise structure of the IR and the allowance of a move to a non-extreme point to allow for a non-extreme point optimal solution. Recenty, Muu and Quy (2003) also proposed a global optimisation algorithm using Branch&Bound and a smoothing approach (see Section 5.4 later on for more discussion on the smoothing approach) for solving the QLPP.

The LBLPP and QBLPP/MPLCC are well structured in some way and as discussed significant progress toward an efficient global optimisation algorithm has been made. For a general case of the nonlinear MPEC/BLPP, the problem becomes more complex due to the unstructured inducible region and ill-condition posed on the problem by the equilibrium constraint (e.g. violation of some constraint qualification and non-differentiability, see Section 5.4.1 for further detail on these issues).

5.3.2 Algorithms for the general BLPP/MPEC

There are mainly four groups of methods proposed for solving the general MPEC including (i) disjunctive nonlinear programming based algorithm, (ii) penalty and interior point based algorithm, (iii) implicit programming based algorithm, and (iv) nonlinear optimisation based algorithm. Each of these groups is discussed in turn as follows.

Disjunctive nonlinear programming based algorithm

One of the most popular ways to reformulate the MPEC as a single level optimisation problem is to replace the *VI* constraint by its Karush-Kuhn-Tucker (KKT) conditions. Naturally, the slackness condition in KKT induces the complementarity constraint (CP) for the single level optimisation problem. This complementarity condition, as mentioned several times, imposes a number of unpleasant features to the single level optimisation problem. Let us define the set of the complementarity constraints as:

$$0 \le g_i(\mathrm{x},\mathrm{y}) \perp h_i(\mathrm{x},\mathrm{y}) \ge 0 \quad \forall i .$$

Then, the complementarity constraints can be classified into one of these sets:

$$\begin{aligned} \alpha &\equiv \{i \mid g_i(\mathrm{x},\mathrm{y}) = 0 \le h_i(\mathrm{x},\mathrm{y})\} \\ \eta &\equiv \{i \mid g_i(\mathrm{x},\mathrm{y}) \ge 0 = h_i(\mathrm{x},\mathrm{y})\} \end{aligned} .$$

The first set is those pairs of the CP with the value of $g_i(\mathrm{x},\mathrm{y})$ at a given (x,y) is equal to zero while $h_i(\mathrm{x},\mathrm{y})$ is greater than or equal to zero; the second set is opposite to the first set where the value of $g_i(\mathrm{x},\mathrm{y})$ is greater than or equal to zero while $h_i(\mathrm{x},\mathrm{y})$ is equal to zero. These two sets represent different ways the CP can be satisfied. Each pair of (α,η) represents a branch of the feasible set of the MPEC. Obviously, at a feasible point (x,y) the union set of all branches of the feasible set is a neighbourhood of (x,y) in the original MPEC. The MPEC with a predefined pair of (α,η) is a decomposed MPEC. Hence, by allowing different combinations of (α,η) that will yield different set of the feasible set of the decomposed MPEC, the MPEC becomes a disjunctive programming problem.

The decomposed MPEC is a well-defined problem (since the CP is removed) which can be solved by an appropriate optimisation algorithm. A well-known approach for solving a disjunctive programming problem like Branch&Bound can be adopted to solve this form of MPEC. Based on this idea Edmunds and Bard (1991) proposed a Branch&Bound algorithm for solving the MPEC. Luo *et al* (1996) proposed a Piecewise Sequential Quadratic Programming (PSQP) method for solving the MPEC in the form of disjunctive programming. Let $\mathcal{A}$(x,y) denote the family of index set pairs (α,η) that satisfy the complementarity condition at a given point (x,y). At each iteration of the

PSQP, for a given initial solution (x,y) a pair of (α,η) is selected from a A(x,y) and a quadratic approximation of the decomposed MPEC is solved to determine the next step of the search. Then, the set of the A at the new (x,y) is generated and the process iterates until no further improving direction can be found from the quadratic approximation problem or the stopping condition is satisfied. This descent search direction of the PSQP is one step of the full SQP method (solving the quadratic approximation problem). The advantage of this method over the Branch&Bound algorithm is that only one step of the full SQP is solved for each branch of the feasible set selected whereas in the Branch&Bound algorithm a full SQP must be used to solve the decomposed MPEC to obtain the bound.

Penalty and interior point based algorithm

A classical way to deal with a constrained nonlinear optimisation is to create an auxiliary function that is a combination of the objective function and the penalised constraint:

$$\theta = f(x) + \mu \cdot (g(x))^2$$

where $f(x)$ is the original objective function of the minimisation problem, $g(x)$ is the equality constraint, and μ is the penalty parameter. Certain methods in the class of penalty based approach from the nonlinear optimisation include quadratic penalty method, log-barrier method, and augmented lagrangian method (Nocedal and Wright, 1999). The main operation for this kind of method is to define a sequence of the penalty parameters. By making this coefficient larger and larger, the violation of the constraint is penalised more and more severely, hence forcing the minimiser of the penalty function closer and closer to the feasible region of the constrained problem. Alternatively, the log-barrier function modifies the objective function surface such that the level of the auxiliary function becomes very high (due to the value of the natural logarithm near zero) near the boundary of the constraints. Hence, the search direction will be always kept inside the feasible region. With this behaviour, the log-barrier method is sometimes referred to as the interior point algorithm.

The other possible method is the exact penalty function in which with a certain choice of μ only one unconstrained optimisation with the auxiliary function can be used to find the optimal solution of the original constrained problem. On the other hand, the inexact penalty methods like quadratic penalty, log-barrier, or augmented lagrangian may need to solve a series of unconstrained problems to achieve the optimal solution.

The application of this type of optimisation method to the MPEC has been a long development. The penalty interior-point algorithm (PIPA) developed by Luo *et al* (1996) is the first penalty based method for MPEC. The complementarity condition of the MPEC (with the transformation of *VI* to its KKT condition) is replaced by the penalised function as explained earlier. A quadratic approximation problem is formed and solved in which the solution of this sub-problem determines the descent direction of the problem. The step-length is then defined by the Armijo inexact line search. However, Leyffer (2002) demonstrated the case where the iteration of PIPA may collapse and the algorithm may not converge to a local optimum of the MPEC.

Benson *et al* (2004) analysed the problems with an interior-point algorithm when applied to MPEC and proposed some heuristic implementation to avoid the difficulty with the MPEC. Similarly, Liu and Sun (2004) proposed the log-barrier penalty based method to solve the MPEC. They also incorporated some treatment with the complementarity condition by relaxing the equality constraint.

Marcotte and Zhu (1997) proposed the exact penalty approach for the MPEC. The equilibrium condition is replaced by the gap function in which the solution algorithm needs to employ the subgradient information due to the non-differentiability of the gap function. As noted, with the proposed method only one unconstrained optimisation could be solved with the exact penalty which is a contrast to other methods proposed. Similarly, Shimizu *et al* (1997) proposed to replace the lower level problem by its optimal value of the lower level. Then, a penalty based method can be used to solve the problem with the information on the directional derivative of the value function (note that in the case of single unique solution of the lower level the value function is differentiable). As discussed earlier, this method has been recently applied to the transport NDP by Meng *et al* (2001).

Implicit programming based algorithm

The MPEC can also be reduced to a single level optimisation problem as follows:

$$\begin{aligned} &\min f\left(x, y(x)\right) \\ &s.t. \\ &x \in X \end{aligned} \qquad \text{Equation 5-1}$$

where $y(x)$ is a solution of the parametric *VI* given *x*. This problem becomes a single level optimisation problem parameterized only by *x*.

This formulation is regularly referred to as the implicit program. One of the key assumptions for the applicability of the implicit program is the uniquely determined solution of the parametric *VI* given *x* from the upper level. This can be satisfied with the strongly monotone condition of the mapping function of the VI (see Theorem 4-1 in the previous chapter).

The root of this idea comes from the seminal work by Fiacco and McCormick (1968) who analysed the KKT condition of a parametric nonlinear program. They utilised the implicit function theorem (Ortega and Rheinboldt, 1970) to define the existence of a locally unique parametric function of the solution to the parametric nonlinear problem (as a function of the parametric term, e.g. *y(x)*). This result opened up the research in the area of sensitivity and stability analysis of a nonlinear program.

Indeed, in solving the problem stated in Equation 5-1 one needs to understand the sensitivity and stability of the parametric *VI* in order to appropriately define the function *y(x)* and its derivative or strictly speaking its directional derivative. The implicit function *y(x)*, which is the solution of the parametric *VI*, is not differentiable (Luo *et al*, 1996). The key result obtained so far is that the implicit function, *y(x)*, is a local Lipschitz continuous function when VI gives a unique solution given *x*. Thus, the directional derivative (or subgradient) of *y(x)* can be defined. In addition, if *f(x)* is a continuously differentiable function, then *f(x,y(x))* is also a locall Lipschitz and directionally differentiable function. With this possibility and the guaranteed Lipschitz condition, the theory of nonsmooth optimisation (see Clarke, 1983) has been adopted as the main tool for developing optimisation algorithms for solving the problem in

Equation 5-1. Pang *et al* (1991) proposed a descent direction search method for solving the non-differentiable implicit program. Also the Implicit Programming Algorithm was proposed in Luo *et al* (1996) and tested extensively in Lim (2002). Similarly, Outrata and Zowe (1995) implemented the descent search algorithm using the directional derivative of the objective function of the implicit program. Some researchers have already attempted to apply this class of technique to the problem in a transport context. Patriksson and Rockafella (2002) applied the Minty parameterisation to find the sensitivity information of the parametric equilibrium condition defined in a Normal cone form (this also relies heavily on the nonsmooth optimisation theory).

The other class of methods widely developed for solving the MPEC problem with the utilisation of the sensitivity information (or directional derivation) of the parametric *VI* is the so called *'bundle method'*. During the iteration process the bundle method constructs and updates the piecewise affine local approximated models of the objective function. These approximated local models are based on the objective values and subgradient at the single iteration points. All bundle methods have two distinct features:

- the subgradient information collected from previous iterations is used to define a better approximation of the model in order to computer a better solution for the current iteration
- if the current approximated model is not good enough, more subgradient information is collected to produce a better approximated problem

Several methods based on the idea of the bundle method have been proposed (see for example Kolstad and Lasdon, 1990; Falk and Liu, 1992; Dempe, 2002).

Nonlinear optimisation based algorithm

The setup of the MPEC does naturally discourage the applications of some existing nonlinear optimisation algorithms due to the failure of MPEC to satisfy the constraint qualification required for those algorithms. Thus, algorithms based on the linearisation of the feasible set may fail because feasibility of the linearisation set may not be guaranteed near the solution (Luo *et al*, 1996). This is indeed the inherit property of the equilibrium constraint and will be discussed in detail in the next section. However, some researchers have attempted to apply existing nonlinear optimisation algorithm to the MPEC. Bard (1988) experimented with application of the gradient projection method to some BLPPs and reported failure on 50-70% of the tests. Conn *et al* (1996) and Ferris and Pang (1997) independently tested the well-known nonlinear optimisation solver, lancelot, and also reported the same failure of the algorithm in solving the MPEC.

These failures from various tests undoubtedly forced the research community to develop a number of specialised methods for solving the MPEC as discussed previously in this section. Nevertheless, there have been some encouraging results from the application of a particular type of nonlinear optimisation algorithm to the MPEC. Fletcher and Leyffer (2002) reported very remarkable results on a large collection of MPEC test problems when applying the Sequential Quadratic Programming (SQP) to the problems. They solved over 100 MPECs and only two cases failed. This encouraging result stimulated a more formal analysis of the specific property of the SQP algorithm that enables it to handle the MPEC and its underlying ill-posed structure.

Following this result Fletcher *et al* (2002) analysed the convergence of the SQP method for MPEC from the theoretical point of view. The key conclusion is a number of conditions of MPEC in which the SQP will be able to solve MPEC successfully including the assumption that all quadratic programming approximations remain consistent during the optimisation process. In addition, they developed a restoration phase (in the filter-SQP solver) that relaxes the complementarity condition when it becomes inconsistent to ensure the convergence of the algorithm near an optimal solution.

Similarly, Anitescu (2000) analysed the convergence of the SQP algorithm for MPEC. In particular, he proposed the elastic mode, relaxing constraint linearisation if they are inconsistent, of the SQP and stressed the difference between his approach and other smoothing approaches which will be discussed next. An alternative approach to MPEC via a nonlinear optimisation algorithm is considered by Andreani and Martinez (2001). Their theoretical construction of the convergence analysis is largely based on an assumption that the MPEC satisfies strict complementarity and a certain optimality condition of the MPEC, the AGP (Approximated Gradient Projection) optimality condition.

The other direction of the attempt to apply a nonlinear optimisation algorithm to the MPEC is to relax or smooth the complementarity constraint (also referred to as the perturbed equilibrium condition) prior to the optimisation process. The relaxed/smoothed MPEC is denoted MPEC(ρ) where ρ is the smoothing parameter. This process is similar to the restoration phase or elastic mode inside the SQP algorithm mentioned earlier. The key difference is that with the relaxed or smoothed MPEC a number of MPEC(ρ) is solved while ρ approach zero. On the other hand, the restoration and elastic mode are only active at some iteration of the optimisation algorithms hence the lower number of optimisation iterations.

Facchinei *et al* (1999) pioneered the smoothing approach to MPEC in which the complementarity constraint is replaced by an equivalent smoothed merit function which is differentiable everywhere. The smoothing parameter represents the level of relaxation of the complementarity constraint (equilibrium condition). When the smoothed parameter for the merit function vanishes the merit function is equivalent to the complementarity constraint. Fukushima *et al* (1998) specialised the smoothing approach for the MPEC with linear complementarity constraint (MPLCC) using the perturbed Fischer-Burmeister function (Fischer, 1995), see detail in the next section. Both methods apply the SQP as the main nonlinear optimisation algorithm for the smoothed problem. The smoothing algorithm is a relatively new idea even in the context of MPEC. The next section is devoted to experiment with the application of the smoothing algorithm to the COTP.

The other form of relaxation was recently proposed by Lin and Fukushima (2003). Instead of relaxing the complementarity condition resulting from the KKT condition of the *VI*, they implemented the relaxation scheme on the variational inequality constraint. Similar to the cutting plane algorithm proposed by Lawphongpanich and Hearn (2004), the variational inequality is formed based on the extreme points of the feasible region. However, the key difference between this method and the cutting plane method is that the feasible region of this method is defined as a positive cone. The feasible region of the actual variable is defined separately as a system of equations. With the feasible region of the *VI* being the positive cone, the extreme points can be defined easily (a vector with the same length as the variable vector but with only one non-zero element).

Not all extreme points are introduced into the VI condition at the same stage. Instead a similar scheme as the cutting plane algorithm is adopted where an additional extreme point is inserted into the *VI* condition as necessary in each iteration. Again, the relaxed problem in each iteration is solved by the SQP algorithm.

5.4 SMOOTHING ALGORITHM: MERIT FUNCTION APPROACH

The review of the optimisation algorithms for MPEC in the previous section opens a new possibility of tackling the NDP and COTP. The purpose of this section is to illustrate the applicability of one of the possible methods which has not been applied to the COTP in the transport context, the smoothing method for MPEC as suggested by Fukushima *et al* (1998) and Facchinei *et al* (1999). The purpose of the smoothing method is to replace the complementarity condition by a well-behaved function (merit function) and then apply some off-shelf optimisation algorithm to the problem. The next section will explain the possible problem with the complementarity constraint. Then, Section 5.4.2 proposes an approach to smooth the complementarity condition to a continuously differentiable function. Section 5.4 shows a way to utilise the smoothing approach for solving the COTP (or the NDP) with multicommodity flow formulation and path-based formulation. Then, the method is tested with a small network.

5.4.1 Problem with the complementarity constraint

Recall the COTP problem as shown in Equation 4-9 in the previous chapter:

$$\begin{aligned}&\min_{(\tau,\mathbf{v},\mathbf{d})} \psi_1(\tau,\mathbf{v},\mathbf{d})\\ &s.t.\\ &0\le\tau_a\le\varepsilon_a\cdot\overline{\tau}_a \qquad \text{for given } \varepsilon \text{ and } \forall a\in A^{*}\\ &(\mathbf{v},\mathbf{d})\rightarrow sol\big(VI\big(F(\mathbf{v},\mathbf{d},\tau),\Omega(\tau)\big)\big)\end{aligned}$$

The VI for the UE condition at the lower level can then be reformulated as the MCP as shown in Equation 4-3 (page 76) and we obtain:

$$\begin{aligned}
&\min_{(\tau,\mathbf{v},\mathbf{d})} \psi_1(\tau,\mathbf{v},\mathbf{d}) && \\
&s.t. && \\
&0 \le \tau_a \le \varepsilon_a \cdot \bar{\tau}_a && \text{for given } \varepsilon \text{ and } \forall a \in A \\
&v_a = \sum_{\forall k} \sum_{\forall p \in P_k} \delta_{p,a} F_p && \forall a \in A \\
&\left(C_p(\mathbf{F},\tau) - \mu_k\right) \cdot F_p = 0, && \forall p \in P_k; \forall k \in K \\
&C_p(\mathbf{F},\tau) - \mu_k \ge 0, && \forall p \in P_k; \forall k \in K \\
&\left(\mu_k - D_k^{-1}(d_k)\right) \cdot d_k = 0, && \forall p \in P_k; \forall k \in K \\
&\mu_k - D_k^{-1}(d_k) \ge 0, && \forall p \in P_k; \forall k \in K \\
&\sum_{p \in P_k} F_p = d_k, && \forall k \in K \\
&\mathbf{F} \ge 0, \mathbf{D} \ge 0, \mathbf{d} \ge 0 &&
\end{aligned} \qquad \text{Equation 5-2}$$

This formulation does look like a normal non-linear optimisation, but with a closer observation the problem with the complementarity condition can be revealed. Two problems are associated with the complementarity condition, (i) the function is not differentiable everywhere, and (ii) the constraint induced by the *MCP* does not satisfy the constraint qualification required for most of the nonlinear optimisation algorithms. This will be discussed in turn.

Non-differentiability of the complementarity condition

Let us concentrate on the complementarity path of the constraints in Equation 5-2:

$$\left(C_p(\mathbf{F},\tau) - \mu_k\right) \cdot F_p = 0.$$

Let $a = C_p(\mathbf{F},\tau) - \mu_k$ and $b = F_p$, then the complementarity condition (CP) above can be simplified as: $a \cdot b = 0, a \ge 0, b \ge 0$. The graphical illustration of this condition is:

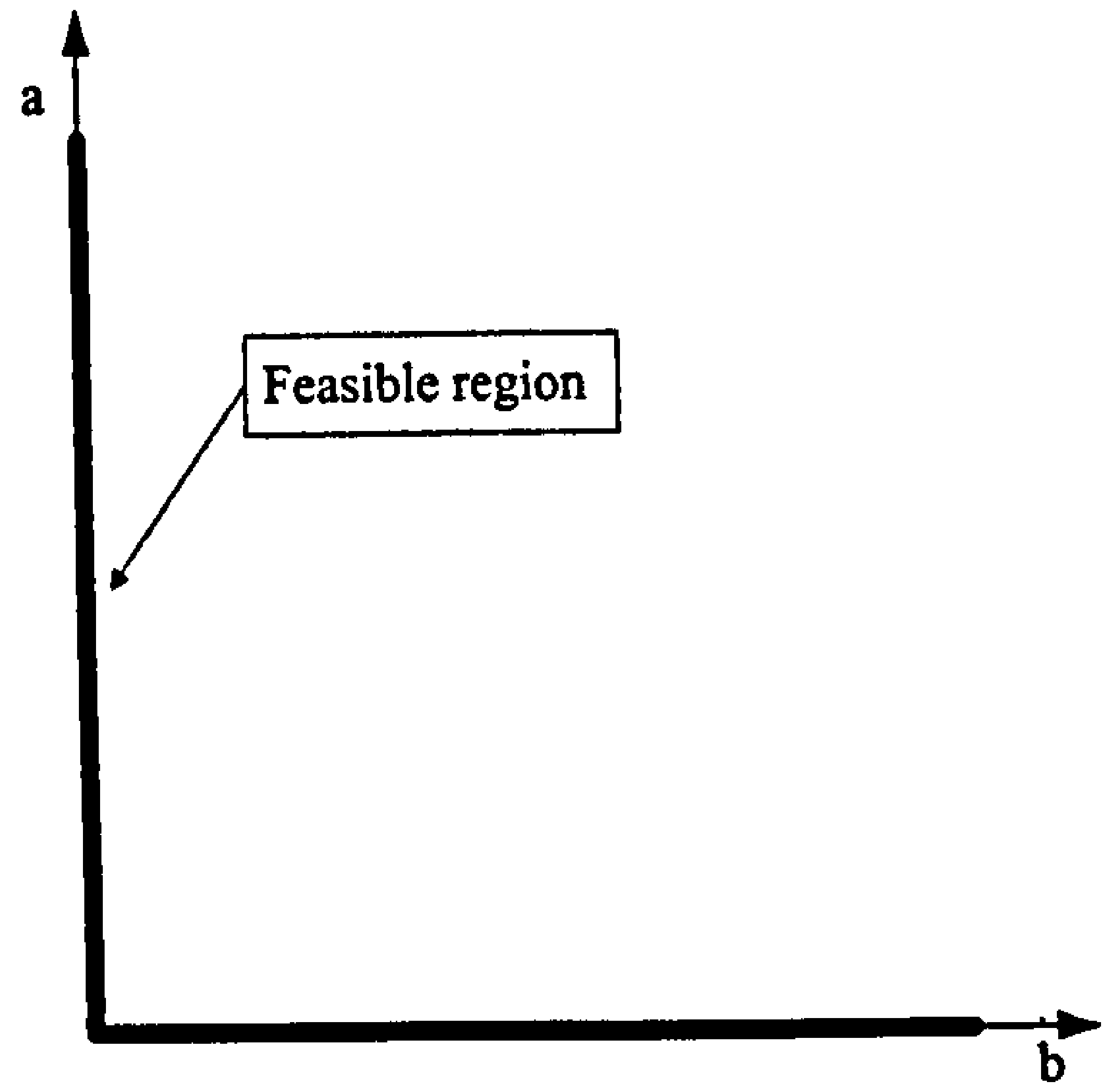


Figure 5-2 Feasible region generated by the complementarity condition

It turns out that at the point of $a = b = 0$ (origin point of Figure 5-2), the CP is non-differentiable. The other way to illustrate this is to replace the CP by the min function:

$$0 \le a \perp b \ge 0 \equiv \min(a,b) = 0; a,b \ge 0,$$

then replace the min function by the Fischer-Burmeister function, F-B function, (Fischer, 1995) where:

$$\min(a,b) \equiv \phi(a,b) = (a+b) - \sqrt{a^2+b^2}.$$

The Jacobian of the F-B function can be written as:

$$\nabla\phi(a,b) \equiv \begin{pmatrix} \dfrac{\partial\phi(a,b)}{\partial a} \\ \dfrac{\partial\phi(a,b)}{\partial b} \end{pmatrix} = \begin{pmatrix} 1 - \dfrac{a}{\sqrt{a^2+b^2}} \\ 1 - \dfrac{b}{\sqrt{a^2+b^2}} \end{pmatrix} \qquad \text{Equation 5-3}$$

As mentioned, from the Jacobian expression above the F-B function is not differentiable at $a = b = 0$ since the denominators of the second terms in the right hand side of Equation 5-3 will become zero. If one implements directly the nonlinear optimisation algorithm to Equation 5-2, the optimisation algorithm may encounter the problem with the non-differentiability of the CP.

Violation of the constraint qualification of the complementarity condition

The second problematic feature of the MCP is the violation of the constraint qualification. Recall the simplified complementarity condition:

$$a \cdot b = 0, a \geq 0, b \geq 0 .$$

If $a = 0$, the gradient that corresponds to this active constraint and the gradient of the complementarity constraint $a \cdot b = 0$ are linearly dependent. On the other hand, if $b = 0$, the gradient associated to the complementarity constraint $a \cdot b = 0$ is dependent of the gradient associated to $b = 0$. Since, at any feasible point of the Equation 5-2, either a = 0 or b = 0, it turns out that the set of gradients of active constraints are linearly dependent for all the feasible points of MPEC. Example 5-1 illustrates the linear dependent problem with MPEC.

Example 5-1 Linear dependent problem with CP in MPEC

Consider a small example due to Jiang and Ralph (1997):

$$\begin{aligned}
&\min_{(z_1,z_2)} (z_1 - 1)^2 + z_2^1 \\
&s.t. \\
&z_2 \geq 0 \\
&(z_2 - z_1) \geq 0 \\
&z_2 \cdot (z_2 - z_1) = 0
\end{aligned}$$

Figure 5-3 shows the contour of the objective function (broken line) and the feasible region of the problem (bold line). From the graphical illustration of this problem, the optimal solution to this problem is $z_1 = z_2 = \frac{1}{2}$. Now consider the first-order condition of the Lagrangian of the problem at the optimal solution ($z_1 = z_2 = \frac{1}{2}$):

$$\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \lambda \cdot \begin{pmatrix} -1 \\ 1 \end{pmatrix} - \gamma \cdot \begin{pmatrix} -1/2 \\ 1/2 \end{pmatrix},$$

where λ and γ are the Lagrange multipliers associated with the constraints $(z_2 - z_1) \geq 0$ and $z_2 \cdot (z_2 - z_1) = 0$ respectively (thus λ is restricted to non-negative value and γ is free in sign). At the optimal solution $z_2 = \frac{1}{2} \geq 0$. Thus, the constraint $z_2 \geq 0$ is not binding and does not need to be included into the first-order optimality condition above.

Obviously, from the optimality condition shown the gradient of the active constraint are linearly dependent. The effect of this situation is that the Lagrange multipliers for the active constraints become unbounded which can be defined by:

$$\varpi = \left\{(\lambda,\gamma)\middle|\lambda \geq 0, \lambda - \frac{1}{2}\cdot\gamma = 1\right\},$$

which can be depicted as shown in Figure 5-4.

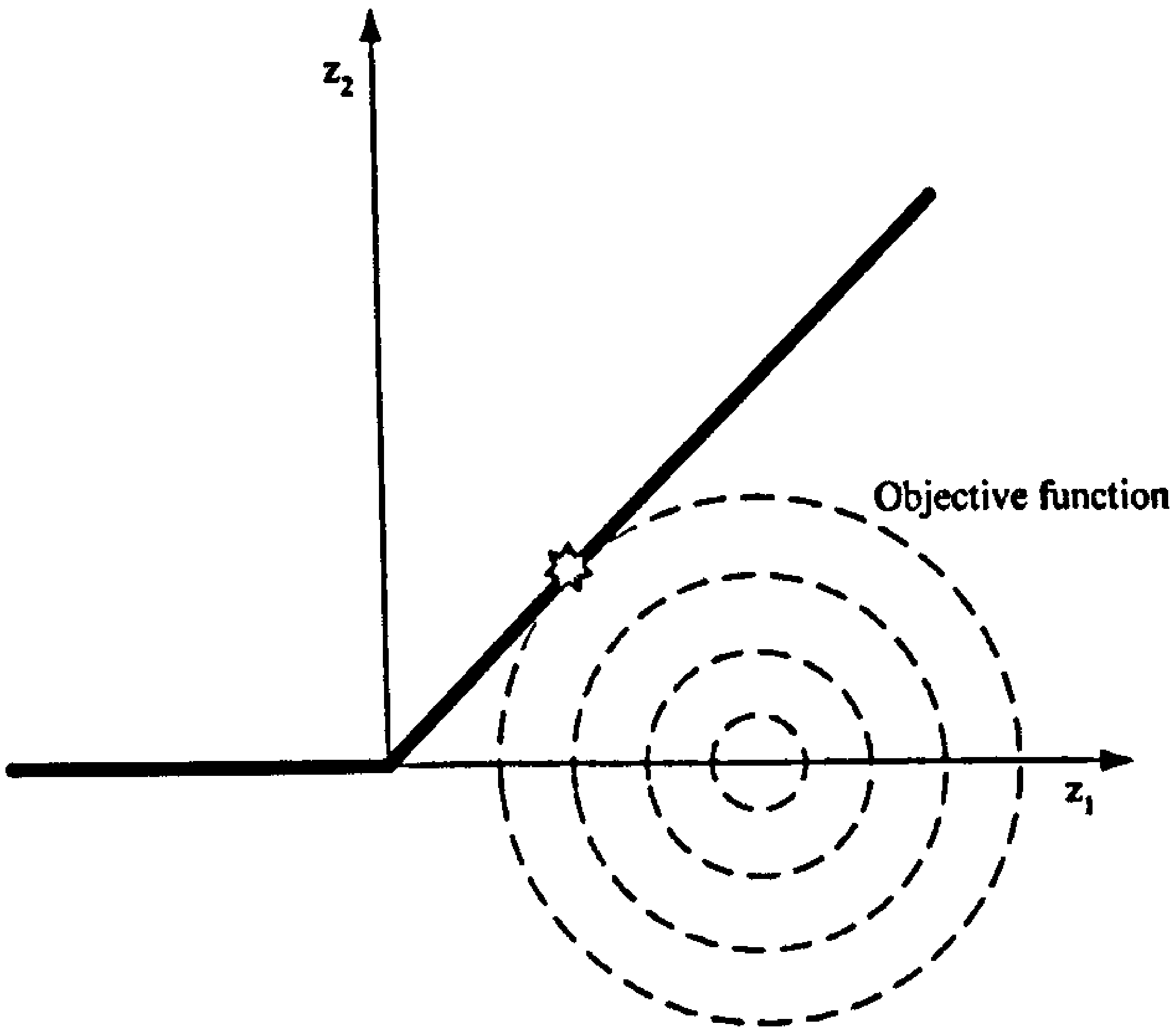


Figure 5-3 Illustrative example of the linear dependent problem with MPEC

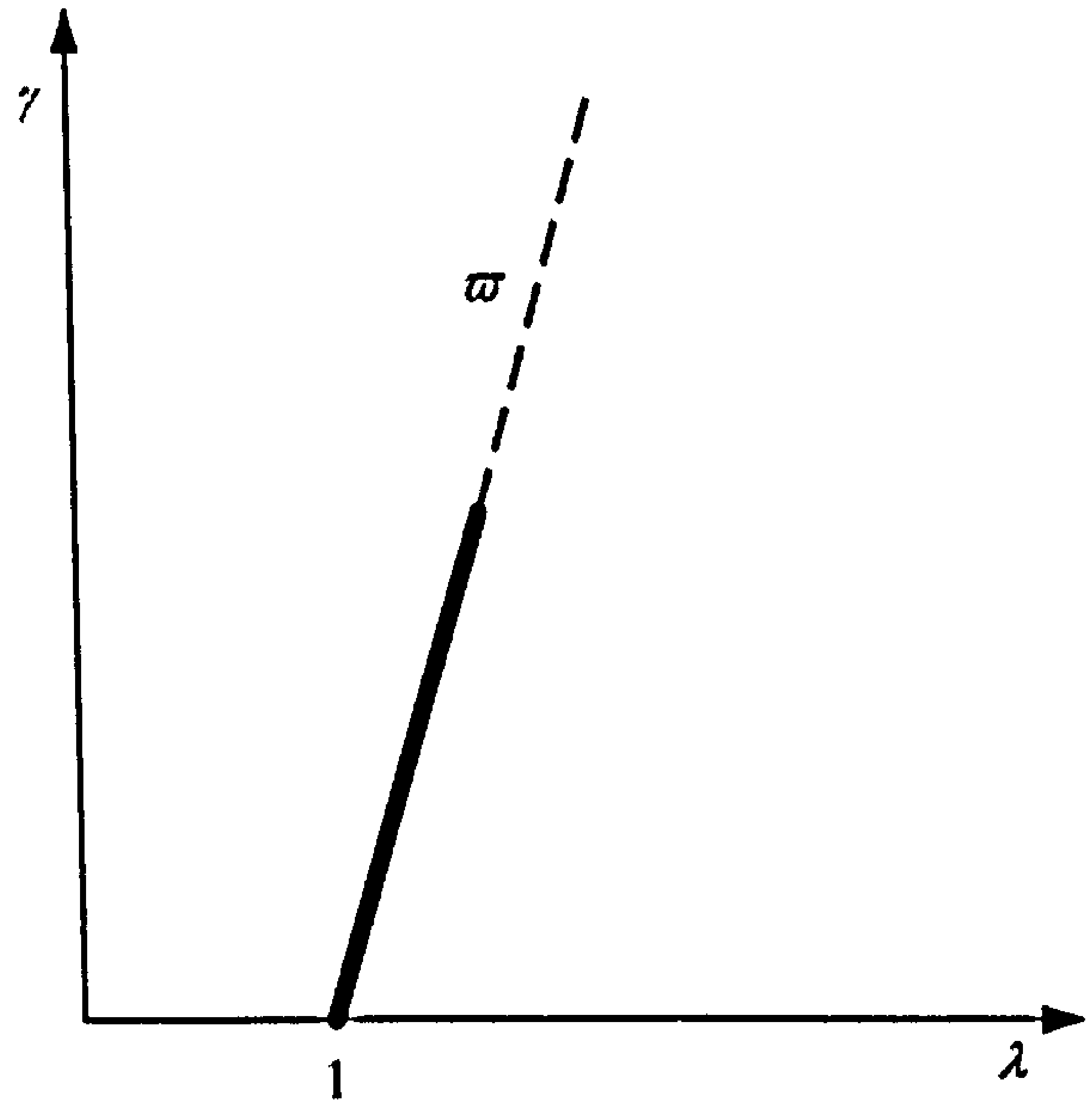


Figure 5-4 Multiplier sets of the illustrative example

This unbounded set of multipliers imposes a problem with applying most of the existing nonlinear optimisation algorithms.

Example 5-1 clearly shows the possible problem with the CP. Indeed, the fact that all feasible point of the MPEC violate the linear independent constraint qualification implied that it also violated the Magasarian Fromovitz Constraint Qualification (MFCQ) at every feasible point, see Chen and Florian (1995) and Scheel and Scholtes (2000). Since MFCQ is a sufficient condition for stability of a nonlinear program, the lack of MFCQ has been advocated as a theoretical argument against the use of standard nonlinear optimisation algorithm. A number of numerical tests of applying standard nonlinear optimisation algorithms have reported their failures(see Bard, 1988; Conn *et al*, 1996; Ferris and Pang, 1997).

5.4.2 Smoothing the complementarity constraint

With the two stringent problematic features of the MPEC discussed in the previous section, this section proposes a smoothing method that aims to enable the application of a standard nonlinear optimisation algorithm with the MPEC.

Recall the reformulated CP as the F-B function:

$$\min(a,b) \equiv \phi(a,b) = (a+b) - \sqrt{a^2+b^2}.$$

As shown in the previous section, the F-B function and its corresponding CP condition is not differentiable everywhere. The strategy to avoid the non-differentiable condition is to introduce the perturbed F-B function:

$$\phi_\rho(a,b) = (a+b) - \sqrt{a^2+b^2+\rho}.$$

The key property of this perturbed or smoothed F-B function is that:

$$\lim_{\rho \to 0} \phi_\rho(a,b) = \min(a,b),$$

and ϕ_ρ is continuously differentiable everywhere when $\rho \neq 0$. The function $\phi_\rho(a,b)$ is therefore a smooth perturbation of the complementarity condition. Figure 5-5 and Figure 5-6 compare the contour of the smoothed F-B and the min functions (observe the non-smoothness of the min function and the smoothing effect of the perturbed F-B function).

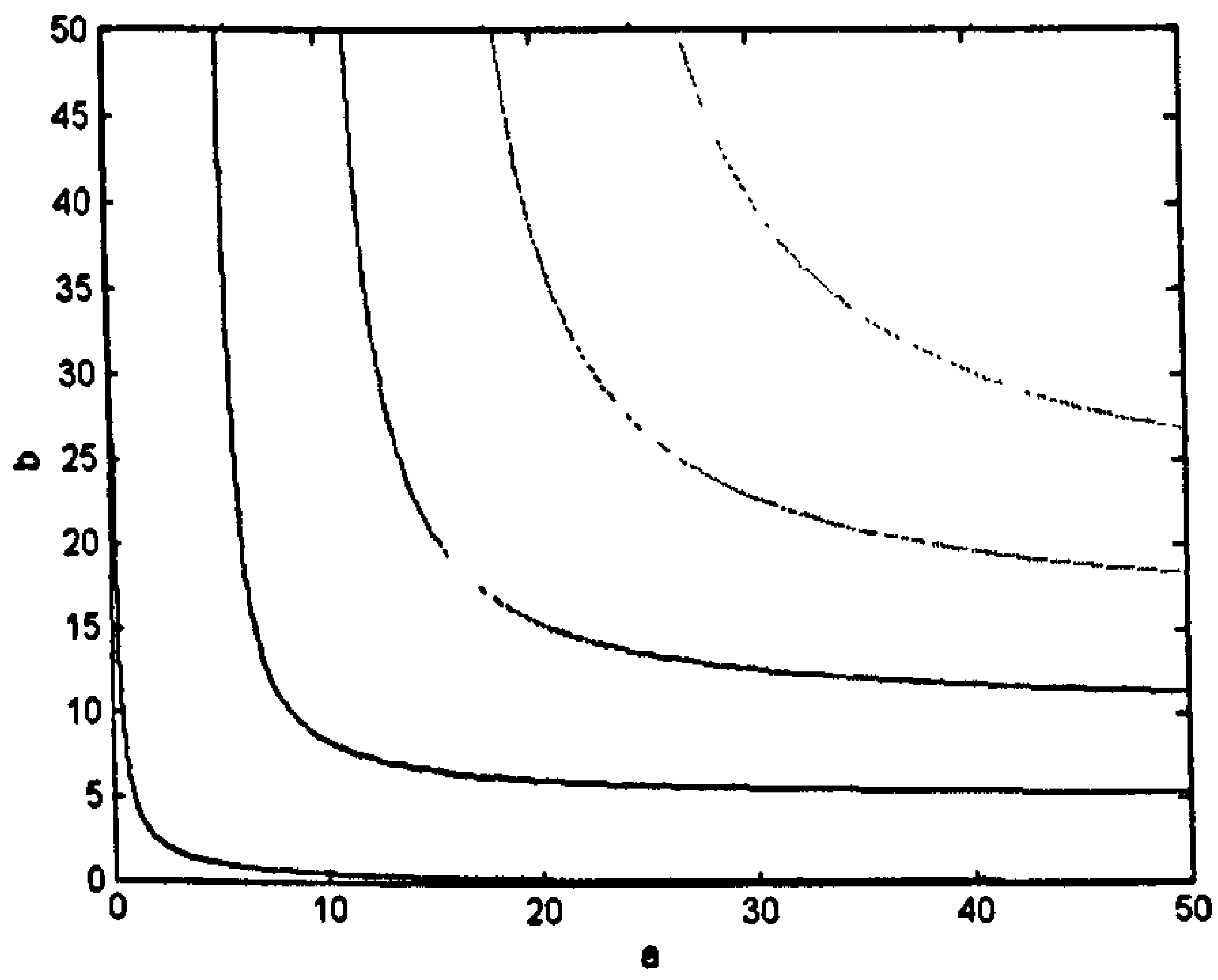


Figure 5-5 Contours of the smoothed F-B function (with $\rho = 10$)

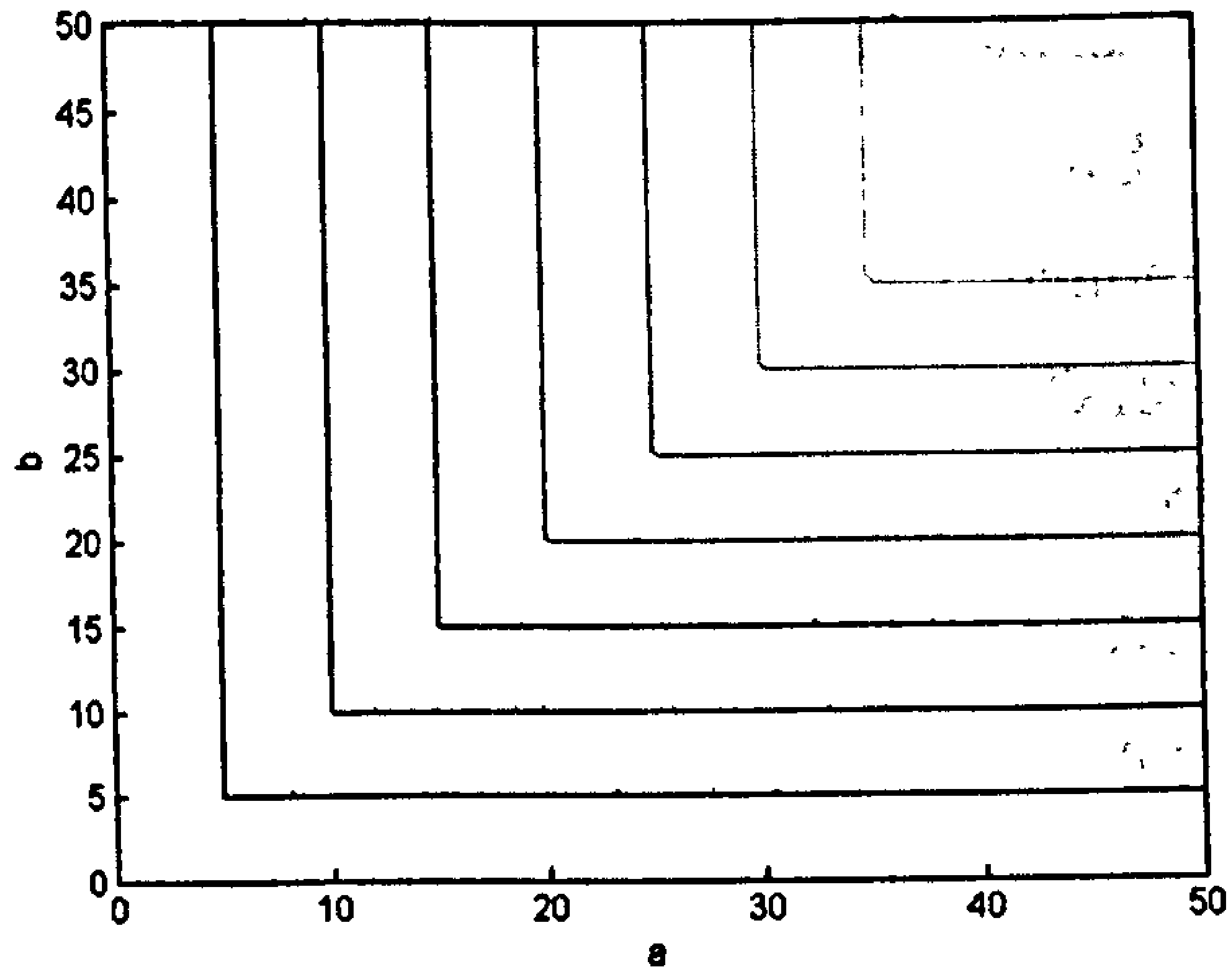


Figure 5-6 Contours of the min function

There exist other possible forms of the smoothing function for the min function. For instance, Scholtes and Stöhr (1999) proposed a piecewise smooth function for the complementarity condition and applied it with the exact penalty based approach. Kanzow (1996) proposed a different form of the smoothing function.

In parallel, the introduction of the smoothed F-B function also eliminates the linear dependence problem discussed in the previous section. This can be illustrated, again, in Figure 5-5 and Figure 5-6. For the min function, the linearly dependence condition is illustrated clearly. On the other hand, the illustration in Figure 5-5 illustrates the effect of the perturbed parameter that also eliminate the linear dependence problem of the active constraints. From the Jacobian of the original F-B function in Equation 5-3, the Jacobian of the smoothed F-B function can be redefined as:

$$\nabla\phi(a,b) \equiv \begin{pmatrix} \dfrac{\partial\phi(a,b)}{\partial a} \\ \dfrac{\partial\phi(a,b)}{\partial b} \end{pmatrix} = \begin{pmatrix} 1-\dfrac{a}{\sqrt{a^2+b^2+\rho}} \\ 1-\dfrac{b}{\sqrt{a^2+b^2+\rho}} \end{pmatrix} \qquad \text{Equation 5-4}$$

Notice that with the original F-B function, the problem of the linear dependency of the active constraints is still valid. However, with the perturbed F-B function when $\rho \neq 0$ the Jacobian of the perturbed F-B function shown above does not suffer from the linear dependence problem with its associated active constraint (in this case either with $a \geq 0$ or $b \geq 0$).

With an available well-behaved function representing the complementarity function, we are ready to state the smoothing algorithm for solving the MPEC. Let $\mu_k = D_k^{-1}(d_k)$, we can then define the smoothed MPEC problem by replacing the complementarity condition with the perturbed F-B function:

$$\begin{array}{ll}
\min\limits_{(\tau,\mathbf{F},\mathbf{d})} \psi_1(\tau,\mathbf{F},\mathbf{d}) & \\
s.t. & \\
0 \leq \tau_a \leq \varepsilon_a \cdot \bar{\tau}_a & \text{for given } \varepsilon \text{ and } \forall a \in A \\
\phi_\rho\left(\left(C_p(\mathbf{F},\tau) - D_k^{-1}(d_k)\right), F_p\right) = 0, & \forall p \in P_k; \forall k \in K \\
C_p(\mathbf{F},\tau) - D_k^{-1}(d_k) \geq 0, & \forall p \in P_k; \forall k \in K \\
\sum\limits_{p \in P_k} F_p = d_k, & \forall k \in K \\
\mathbf{F} \geq \mathbf{0}, \mathbf{D} \geq \mathbf{0}, \mathbf{d} \geq \mathbf{0} &
\end{array}$$

Equation 5-5

The problem stated in Equation 5-5 is named COTP$_\rho$. The outline of the algorithm for solving the COTP is as follows:

Algorithm 5-1:

Step 1: Set ρ^l to be a constant value that is greater than zero and set l = 1. Set $(\tau^0,\mathbf{F}^0,\mathbf{d}^0)$ to be any feasible vectors; pre-specify $0 < \alpha < 1$ and maximum number of iterations.

Step 2: Solve COTP$_{\rho^l}$ by a non-linear optimisation algorithm with the starting point $(\tau^{l-1},\mathbf{F}^{l-1},\mathbf{d}^{l-1})$ and obtain $(\tau^l,\mathbf{F}^l,\mathbf{d}^l)$ as the solutions of COTP$_{\rho^l}$

Step 3: Set l = l+1 and update $\rho^{l+1} = \alpha \cdot \rho^l$; where $0 < \alpha < 1$ is a pre-specified constant. Then, go to Step 2 until reaching the maximum iteration numbers.

The nonlinear optimisation algorithm adopted in this chapter is the Sequential Quadratic Programming (SQP) algorithm (Gill *et al*, 1997).

The problem formulation above is based on the path-flow formulation. An alternative to this formulation is to use the multicommodity link flow based formulation in which the flow conservation constraints are based on the node-link incidence matrix. This formulation is referred to as the node-link formulation. Recall the MCP formulation of the UE condition using the multicommodity link flow defined in Equation 4-4 (page 76). The only modification made is to replace the path-flow complementarity condition in Equation 5-2 with the multicommodity link flow complementarity condition from Equation 4-4. Then, the smoothing F-B function can be applied to the complementarity part of the MCP:

$$x_{i,j}^{k} \cdot \left[c_a \left(v_a, \tau_a \right) - \left(\lambda_j^k - \lambda_i^k \right) \right] = 0 \qquad \forall (i,j) \in A; \forall k \in K .$$

As mentioned, an advantage of the multicommodity link flow based formulation is that it is a complete formulation based on the existing structure of the network. This can be formed very easily with the node-link incidence matrix. On the other hand, the path-based formulation does not allow a direct set-up since identifying all paths in the network can be a difficult task (or some kind of iterative process has to be employed which reduces the completeness of the initial formulation). However, as discussed in the last chapter, the number of variables in the path-based formulation can be substantially smaller than that from the multicommodity link flow based formulation. Next the smoothing algorithm is tested with a small network.

5.4.3 Tests of the merit-function approach with small network

This section presents some numerical experiments with a five link network under fixed demand condition with two O-D pairs between nodes 1-4 and 3-4 (see Figure 5-7). The objective function considered here is to maximise the revenue.

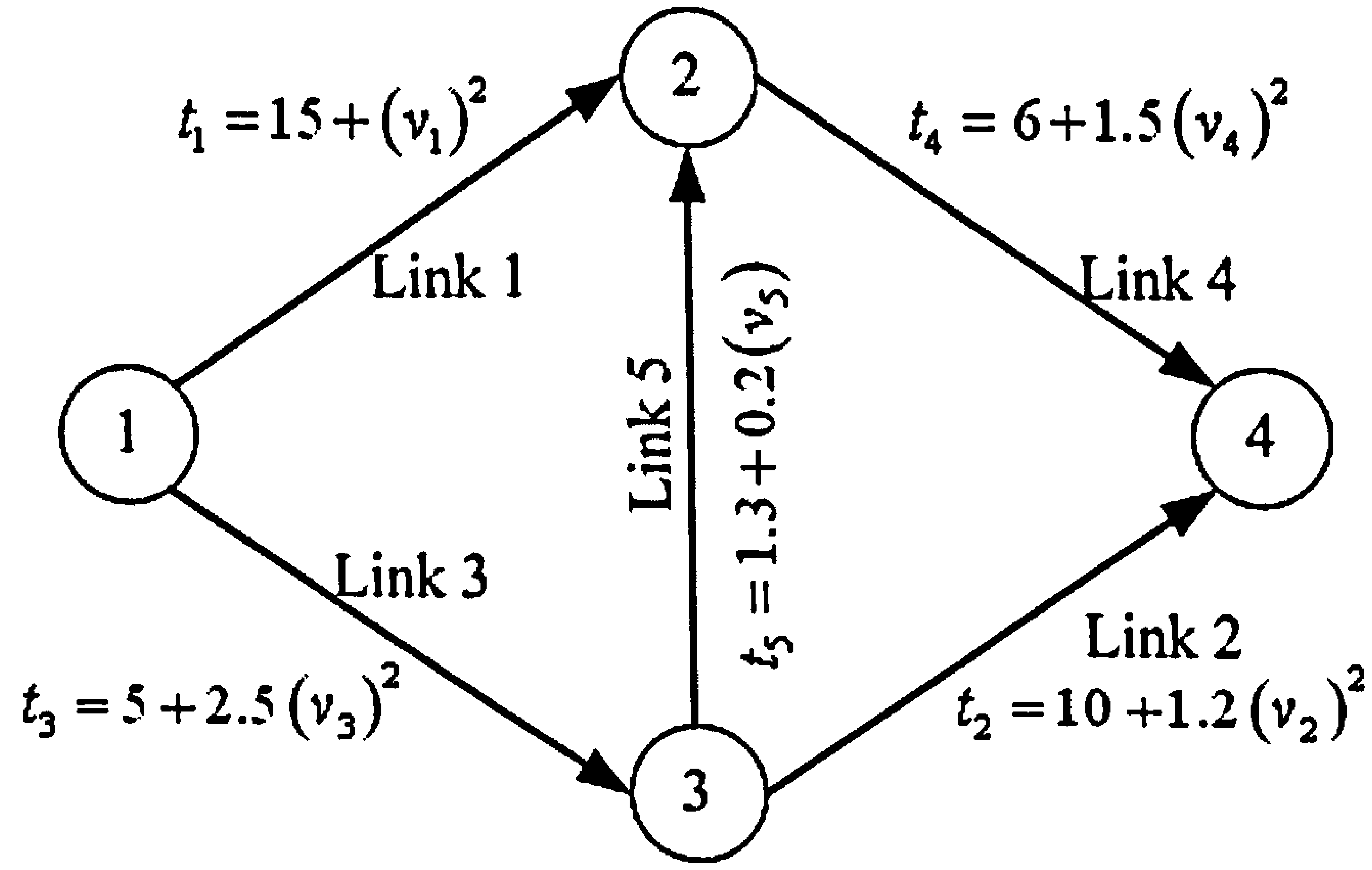


Figure 5-7 Five-link network used for testing the smoothing algorithm for solving COTP

Before discussing the result, let define the paths on the five-link network:

- path 1: link 1 → link4
- path 2: link 3 → link 2
- path 3: link3 → link5 → link4
- path 4: link2
- path 5: link5 → link4

Obviously, paths 1, 2, and 3 connect between the first OD pair (node 1 to node 4) and paths 4 and 5 connect between the second OD pair (node 3 to node 4).

The complementarity condition for the equilibrium equilibrium condition using path flows can be formulated as follows:

$$\begin{aligned}
&\left(C_1(\mathbf{F},\tau)-\mu_1\right)\cdot F_1=0\\
&\left(C_2(\mathbf{F},\tau)-\mu_1\right)\cdot F_2=0\\
&\left(C_3(\mathbf{F},\tau)-\mu_1\right)\cdot F_3=0\\
&\left(C_4(\mathbf{F},\tau)-\mu_2\right)\cdot F_4=0\\
&\left(C_5(\mathbf{F},\tau)-\mu_2\right)\cdot F_5=0
\end{aligned}$$

These complementarity conditions are then replaced by the smoothing function where $a=\left(C_p(\mathbf{F},\tau)-\mu_k\right)$ and $b=F_p$. Thus, the the equilibrium conditions for the optimal toll problem in the form of smoothing function are:

$$\begin{aligned}
&\left(C_1(\mathbf{F},\tau)-\mu_1\right)+F_1-\sqrt{\left(C_1(\mathbf{F},\tau)-\mu_1\right)^2+\left(F_1\right)^2+\rho}=0\\
&\left(C_2(\mathbf{F},\tau)-\mu_1\right)+F_2-\sqrt{\left(C_2(\mathbf{F},\tau)-\mu_1\right)^2+\left(F_2\right)^2+\rho}=0\\
&\left(C_3(\mathbf{F},\tau)-\mu_1\right)+F_3-\sqrt{\left(C_3(\mathbf{F},\tau)-\mu_1\right)^2+\left(F_3\right)^2+\rho}=0\\
&\left(C_4(\mathbf{F},\tau)-\mu_2\right)+F_4-\sqrt{\left(C_4(\mathbf{F},\tau)-\mu_2\right)^2+\left(F_4\right)^2+\rho}=0\\
&\left(C_5(\mathbf{F},\tau)-\mu_2\right)+F_5-\sqrt{\left(C_5(\mathbf{F},\tau)-\mu_2\right)^2+\left(F_5\right)^2+\rho}=0
\end{aligned}$$

These conditions are used in the bilevel optimal toll problem with other necessary constraints.

The toll on each link is then optimised in turn. The results are shown in Table 5-1. The aggregated merit function is the sum of all the smoothed F-B functions which represents the relative gap between the real equilibrium solution and the current solution.

link	optimal toll	revenue	aggregated merit function
1	115.69	319.06	-0.010110
2	166.05	643.16	-0.010220
3	110.41	198.91	-0.000005
4	252.16	803.53	-0.000730
5	13.32	7.15	-0.000200

Table 5-1 Test results with single tolled link with the five-link network (smoothing approach)

Figure 5-8 below depicts the revenue curve as a function of the toll on link 1. The figure certifies the optimality of the toll level found by the smoothing algorithm (around 115).

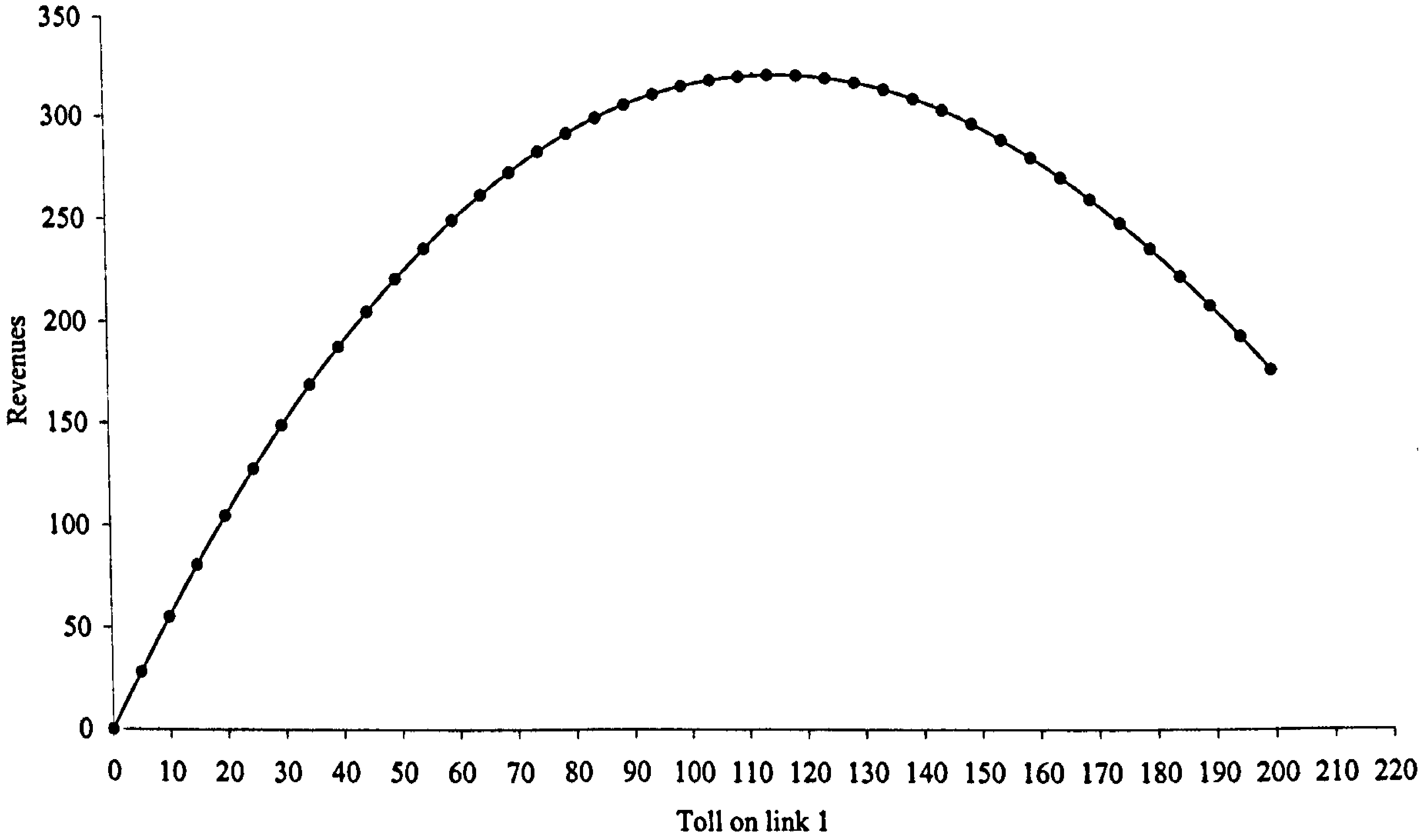


Figure 5-8 Revenue curve as the function of the toll on link 1 (five-link network)

To illustrate the effect of the perturbation parameter (ρ) on the equilibrium condition of the UE, the result with the toll on link 5 is discussed more fully. The sequence of ρ is set to $\{1,0.1,0.01,0.001,0.0001\}$. Note that the column 'Agg. merit' is the aggregated perturbed F-B function and the column 'min. OD cost' is the minimum OD travel cost between the first and second OD pairs. As expected, the aggregated merit function decreased as the value of ρ decreased. Similarly, the deviation of the path costs and their relevant minimum OD travel cost also decreased. In the first iteration (iter.), path 3 had a slightly higher cost than the minimum OD cost of OD1 and similarly path 5 also had a slightly higher cost than the minimum OD cost of OD2. Eventually, all paths for each OD pair have very close path costs in the last iteration.

iter.	ρ	toll	Agg. merit	path cost					min. OD cost	
				path1	path2	path3	path4	path5	OD1	OD2
1	1	13.019	-0.296	130.25	130.28	131.82	89.51	91.04	130.25	89.51
2	0.1	13.016	-0.025	130.11	130.11	130.28	90.16	90.34	130.11	90.16
3	0.01	13.015	-0.003	130.09	130.09	130.11	90.24	90.26	130.09	90.24
4	0.001	12.783	-0.001	130.12	130.12	130.12	90.13	90.13	130.12	90.13

Table 5-2 Iteration results for tolling on link 5 of the five-link network (smoothing approach)

Table 5-3 shows the results of the optimisation result with two tolled links for the five-link network.

link	optimal toll	revenue	aggregated merit function
1,2	53.32, 88.28	736.54	-0.001722
2,3	165.60, 54.00	766.11	-0.000400
1,4	113.81, 138.84	817.64	-0.000007

Table 5-3 Test results with dual tolled links with the five-link network (smoothing approach)

From the numerical experiment, it is found that the smoothing approach has a potential for being applied to a more general problem. However, several drawbacks of this approach were discovered during the tests including:

- the algorithm seems to be very sensitive to the starting point; the algorithm could converge to different solutions with different starting points;
- the initial value and the rate of decrease of the perturbation parameter (ρ); during the tests, it was found that some adjustment is needed to ensure the appropriate value of ρ and its reduction rate; if ρ decreases too rapidly, the problem may become infeasible;
- with the multicommodity link flow formulation the number of variables can be too large especially when we consider a large scale network; on the other hand if the path flow based formulation is used, some form of path generation approach is needed.

A similar approach to this smoothing algorithm was proposed in Clegg *et al* (2001) although they did not present their algorithm as a smoothing approach. The UE condition adopted in their algorithm is represented by the square of the complementarity condition. Furthermore, some form of relaxation is employed where the concept of ϵ-equilibrium is introduced. ϵ-equilibrium can be seen as a form of perturbed equilibrium where a certain amount of tolerance of the complementarity condition is allowed.

5.5 IMPROVED CUTTING PLANE ALGORITHM APPROACH

5.5.1 Cutting plane algorithm for the second-best toll problem

As discussed earlier in Section 5.2, the most recent proposed algorithm for tackling a practical optimal toll problem is the cutting plane algorithm (CPA) (Lawphongpanich and Hearn, 2004). This section explains more detail of the algorithm and tests it with a network.

The CPA defines the UE condition as a variational inequality as discussed in Theorem 4-3 in the previous chapter:

$$\mathbf{c}\left(\mathbf{v}^*\right)^T \cdot \left(\mathbf{v}-\mathbf{v}^*\right) - \mathbf{D}^{-1}\left(\mathbf{d}^*\right)^T \cdot \left(\mathbf{d}-\mathbf{d}^*\right) \geq 0 \text{ for } \forall \left(\mathbf{v},\mathbf{d}\right) \in \Omega .$$

The feasible region of the flow vectors, Ω, is defined by a linear equation system of flow conservation. Thus, Ω is a bounded polyhedral set. From the convex set theory, Ω can be defined by the set of its extreme points (due to the convexity of the feasible region).

Let **H** be the matrix whose columns are the extreme points of Ω, defined by a pair of vector $(\mathbf{u},\mathbf{q})^T$. Then, for any $\left(\mathbf{v},\mathbf{d}\right) \in \Omega$, $\left(\mathbf{v},\mathbf{d}\right)^T = \mathbf{H}\cdot\theta$, for some $\theta \geq 0$, where θ is a column vector and $\sum_i \theta_i = 1$. Basically, this condition implies that $\left(\mathbf{v},\mathbf{d}\right) \in \Omega$ can be defined as a convex combination of a set of extreme point (see Theorem 2.1.6 in Bazaraa *et al*, 1993). Thus, the *VI* for the UE condition can be redefined as the function of the extreme points of Ω:

$$\mathbf{c}\left(\mathbf{v},\tau\right)^T \cdot \left(\mathbf{u}^e-\mathbf{v}\right) - \mathbf{D}^{-1}\left(\mathbf{d}\right)^T \cdot \left(\mathbf{q}^e-\mathbf{d}\right) \geq 0 \text{ for } \forall e \in E$$

where $(\mathbf{u}^e,\mathbf{q}^e)$ is the vector of extreme link flow and demand flow indexed by the superscript e, and E is the set of all extreme points of Ω. Thus, the COTP can be redefined as:

$$\begin{aligned}
&\min_{(\tau,\mathbf{v},\mathbf{d})} \psi_1(\tau,\mathbf{v},\mathbf{d}) \\
&s.t. \\
&0 \le \tau_a \le \varepsilon_a \cdot \bar{\tau}_a \qquad \text{for given } \varepsilon \text{ and } \forall a \in A \\
&(\mathbf{v},\mathbf{d}) \in \Omega \\
&\mathbf{c}(\mathbf{v},\tau)^T \cdot (\mathbf{u}^e - \mathbf{v}) - \mathbf{D}^{-1}(\mathbf{d})^T \cdot (\mathbf{q}^e - \mathbf{d}) \ge 0 \quad \text{for } \forall e \in E
\end{aligned} \qquad \text{Equation 5-6}$$

Indeed, this is the formulation proposed by Lawphongpanich and Hearn (2004). The practical solution algorithm for solving the problem in Equation 5-6 is to sequentially generate and include the necessary extreme points into the set E. The problem stated in Equation 5-6 is referred to as the 'Master Problem'. The key to the algorithm is the sub-problem used to generate the necessary extreme points from given $(\mathbf{v},\mathbf{d},\tau)$ from the Master Problem. The strategy adopted is to include the most rapid descent direction (to achieve the UE condition) for given $(\mathbf{v},\mathbf{d},\tau)$ into the set E. That is to solve the problem:

$$\begin{aligned}
&\min_{(\mathbf{u},\mathbf{t})} \mathbf{c}(\mathbf{v},\tau)^T \cdot \mathbf{u} - \left(\mathbf{D}^{-1}(\mathbf{d})\right)^T \cdot \mathbf{q} \\
&s.t. \\
&(\mathbf{u},\mathbf{t}) \in \Omega
\end{aligned} \qquad \text{Equation 5-7}$$

This problem is referred to as the 'Sub Problem'. One may notice the similarity between the problem in Equation 5-7 and the sub-problem in the Simplical Decomposition Algorithm (or in the Frank-Wolfe algorithm) for solving the normal traffic assignment problem.

Indeed, the problem in Equation 5-7 can be treated in the same manner as the sub-problem in the Simplical Decomposition Algorithm in which the problem can be decomposed into a number of separated problems for different O-D movements and each problem is simply the shortest path problem. Alternatively, the Sub Problem above can be solved as a normal linear optimisation problem with the flow conservation constraint. The CPA can then be summarised as follows:

Algorithm 5-2

```
Step 0: Initialise the problem by finding the shortest paths for each O-D pair; set $l = 0$;
        define the aggregated link flow and demand flow $(\mathbf{u}^l,\mathbf{d}^l)$; and include $(\mathbf{u}^l,\mathbf{d}^l)$ into $E$.
Step 1: $l = l + 1$; Solve the Master Problem with all extreme points in $E$ and get the
        solution vector $(\mathbf{v},\mathbf{d},\tau)$; then set $(\mathbf{v}^l,\mathbf{d}^l,\tau^l)$.
Step 2: Solve the Sub Problem with $(\mathbf{v}^l,\mathbf{d}^l,\tau^l)$ and obtain the new extreme point $(\mathbf{u}^l,\mathbf{d}^l)$;
Step 3: Termination check; if $\mathbf{c}(\mathbf{v}^l,\tau^l)^T \cdot \mathbf{u}^l - (\mathbf{D}^{-1}(\mathbf{d}^l))^T \cdot \mathbf{q}^l \geq 0$, terminate and
        $(\mathbf{v}^l,\mathbf{d}^l,\tau^l)$ is the solution to the COTP, otherwise include $(\mathbf{u}^l,\mathbf{d}^l)$ into $E$ and
        return to Step 1.
```

One of the possible problems with the formulation in Equation 5-6 is that the *VI* constraints may violate the MFCQ (Magasarian Fromovitz Constraint Qualification) as discussed previously in Section 5.4.1. Recall the definition of the MFCQ:

Definition 5-1 (Nocedal and Wright, 1999 page 353) Given an optimisation problem:

$$\begin{aligned} &\min_{\mathbf{x}} f(\mathbf{x}) \\ &s.t. \\ &c_i(\mathbf{x}) = 0 \quad \forall i \in E \\ &c_i(\mathbf{x}) \geq 0 \quad \forall i \in I \end{aligned}$$

, and given the point $\mathbf{x}^*$ and the active set $A(\mathbf{x}^*)$ are those inequality constraints that are binding, we say that the Magararian Fromovitz Constraint Qualification (MFCQ) holds if there exists a vector w such that:

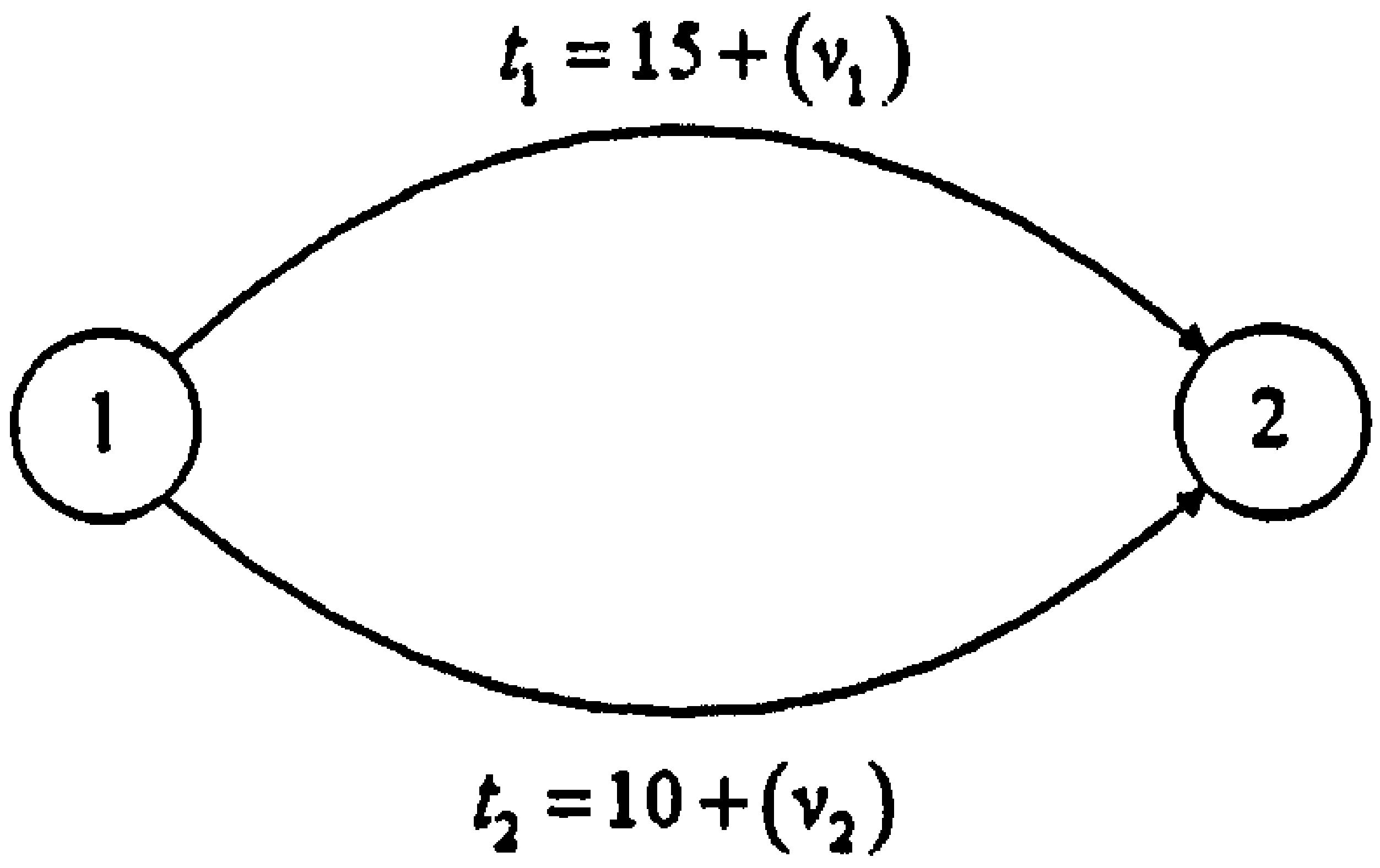
$$\begin{aligned} &\nabla c_i(\mathbf{x}^*)^T \cdot \mathrm{w} > 0 \quad \text{for } \forall i \in A(\mathbf{x}^*), \\ &\nabla c_i(\mathbf{x}^*)^T \cdot \mathrm{w} = 0 \quad \text{for } \forall i \in E, \end{aligned}$$

and the set of equality constraint gradient $\left\{\nabla c_i(\mathbf{x}^*)^T \text{ for } \forall i \in E\right\}$ is linearly independent.

Example 5-2 below following Lawphongpanich and Hearn (2004) illustrates an example of the violation of the MFCQ at every feasible point of Equation 5-6.

Example 5-2 Illustration of the violation of the MFCQ of the COTP with *VI*

Let us consider again the problem stated in Equation 5-6 with only a two-link network below.



For simplicity, let us assume the fixed demand of 20 units from 1 to 2. The two extreme points for this network are $(20,0)^T$ and $(0,20)^T$. A toll level of $0 \le \tau_2 \le 25$ is imposed on link 2 aiming to minimise the total travel time in the network (system optimum). The UE flows on link 1 and link 2 can be defined as:

$$\left(v_1(\tau_2), v_2(\tau_2)\right) = \left[\left(\frac{15+\tau_2}{2}\right), \left(\frac{25-\tau_2}{2}\right)\right].$$

The COTP for this example following Equation 5-6 is:

$$\begin{aligned}
&\min_{(\tau,v)} v_1 \cdot (15 + v_1) + v_2 \cdot (10 + v_2) \\
&s.t. \\
&v_1 + v_2 = 20 \\
&(15 + v_1) \cdot (20 - v_1) + (10 + v_2 + \tau_2) \cdot (0 - v_2) \ge 0 \\
&(15 + v_1) \cdot (0 - v_1) + (10 + v_2 + \tau_2) \cdot (20 - v_2) \ge 0 \\
&0 \le \tau_2 \le 25 \\
&v_1, v_2 \ge 0
\end{aligned} \qquad \text{Equation 5-8}$$

The system optimum solution for this problem is (11.25, 8.75) with the objective function value of 446.875. In addition, the solution to the problem in Equation 5-8 is also (11.25, 8.75, 2.5) with the same objective function as the system optimum, which is not surprising. Let us return to the MFCQ of the problem in Equation 5-8. It is easy to show that the first three constraints are always binding. Based on the gradient of these three binding constraints, the MFCQ requires that there exists a $w \in \Re^3$(each *w* associated with a binding constraint) satisfying the following condition (from Definition 5-1):

$$w_1 + w_2 = 0$$

$$-(\tau_2+10)\cdot w_1 - 35\cdot w_2 - \left(\frac{25-\tau_2}{2}\right)\cdot w_3 > 0$$

$$-\tau_2\cdot w_1 - 15\cdot w_2 + \left(\frac{15+\tau_2}{2}\right)\cdot w_3 > 0$$

Substitute $w_2 = -w_1$ in the two inequalities, the second and third conditions imply that $w_1 > 0.5w_3$ and $w_1 < 0.5w_3$ respectively. Obviously, these two contradictory conditions suggest that MFCQ is not satisfied at any feasible point where $0 \le \tau_2 \le 25$.

The approach to remedy this problem is similar to the method suggested earlier in the smoothing algorithm in which the variational inequality constraint will be relaxed:

$$\mathbf{c}(\mathbf{v},\tau)^T \cdot (\mathbf{u}^e - \mathbf{v}) - \mathbf{D}^{-1}(\mathbf{d})^T \cdot (\mathbf{t}^e - \mathbf{d}) \ge -\rho \quad \text{for } \forall e \in E.$$

Alternatively, in the practical operation of the nonlinear optimisation algorithm, one could set a higher tolerance of the constraint satisfaction to represent the relaxation.

As mentioned, the method of CPA explained above only exploit the polyhedron structure of Ω. The flow conservation constraint as discussed in Section 4.2.1 is still needed in the Master Problem in the multicommodity flow based formulation. This induces a high number of linear constraints and number of variables which is equal to (no. of links × no. of origins) + no. of OD pairs + no. of tolled links. Obviously, in solving a large problem the number of variables and linear constraints can become a problem.

Fortunately, the convex combination of feasible link/demand flows based on a set of extreme points of Ω can also be placed at the upper level:

$$
\begin{aligned}
&\min_{(\tau,\mathbf{v},\mathbf{d})} \psi_1(\tau,\mathbf{v},\mathbf{d}) \\
&s.t. \\
&0 \le \tau_a \le \varepsilon_a \cdot \bar{\tau}_a && \text{for given } \varepsilon \text{ and } \forall a \in A \\
&(\mathbf{v},\mathbf{d})^T = \sum_{\forall e \in E} (\mathbf{u}^e,\mathbf{t}^e)\cdot\theta && \text{Equation 5-9} \\
&\sum_l \theta_l = 1 \\
&\mathbf{c}(\mathbf{v},\tau)^T \cdot (\mathbf{u}^e - \mathbf{v}) - \mathbf{D}^{-1}(\mathbf{d})^T \cdot (\mathbf{t}^e - \mathbf{d}) \ge 0 \quad \text{for } \forall e \in E
\end{aligned}
$$

Notice the replacement of the flow feasibility condition with the convex combination of the extreme points in the second and third constraints. The cutting plane (extreme point of the feasible flow space) generated by the Sub Problem will be inserted into both the VI condition and flow conservation condition. The other process of the algorithm can be proceeded as explained in Algorithm 5-2.

With this improved formulation of the CPA, the number of variables and linear constraints is reduced significantly (the number of variables is now equal to no. of links + no. of OD pairs + no. of tolled links + no. of extreme points). The improved CPA proposed above can also be categorised in the class of 'column generation method' (see Leventhal *et al*, 1973), in which a solution of an approximation of the original problem (in our case Ω) is constructed by replacing the original feasible set with a subset spanned by a finite number of feasible solutions (*E*). The approximation is gradually improved by enlarging the spanning set (*E*) with the generation of a new feasible solution not included in the spanning set previously (generating a shortest path or extreme point of Ω). Similarly, the proposed method can also be categorised as a Simplical Decomposition Algorithm (see Lawphongpanich and Hearn, 1984).

However, it should be noted that this improved cutting plane method is only a 'heuristic' method in which the convergence to a local optimum of the MPEC problem cannot be guaranteed. On the other hand, the original version of the cutting plane algorithm (Lawphongpanich and Hearn, 2004) is guaranteed to converge to a local optimum. The reason for the loss of the local optimum convergence property is the relaxation of the feasible flow in the upper level. The necessary sets of the extreme points for ensuring the optimality conditions of the Master Problem and Sub Problem may not be the same; hence there is no guarantee of the optimality condition of the Master Problem, even if the necessary extreme points for the optimality condition of the Sub Problem are already included. Nevertheless, the experimental results of this method with the Pacman network presented in the next section are very encouraging.

5.5.2 Tests of the improved cutting plane algorithm with Pacman network

In this section, the improved CPA developed in the previous section for solving the COTP is tested with the Pacman network (see Figure 5-9). The same network was also adopted in Shepherd and Sumalee (2004) to illustrate the drawback of the algorithm proposed by Verhoef (2002). The network has 18 links and six OD pairs (from 1 to 5, 5 to 1, 1 to 7, 7 to 1, 5 to 7 and 7 to 5). The network was designed to represent an urban traffic network in which node 1 and node 7 represents the outer zones of the city and node 5 represents the city centre attraction area. The link cost function is defined in the following form:

$$t_j = a_j + b_j \cdot \left(\frac{v_j}{c_j} \right)^{n_j},$$

where the link travel time parameters for the Pacman network are shown in Table 5-4. The elastic demand condition is assumed for this network in which the demand function follows the power law from:

$$d_k = d_k^0 \cdot \left(\frac{\mu_k}{\mu_k^0} \right)^{\beta},$$

where d_k^0 is the demand for OD pair k in the base year, μ_k^0 is the minimum travel cost between OD pair k in the base year, and β is the elasticity value where in this case $\beta < 0$. In a general case, β could adopt different values for different OD pairs. In this test, β is set to be -0.57 for all OD pairs. Table 5-5 shows the values of d_k^0 and μ_k^0 for each OD pair.

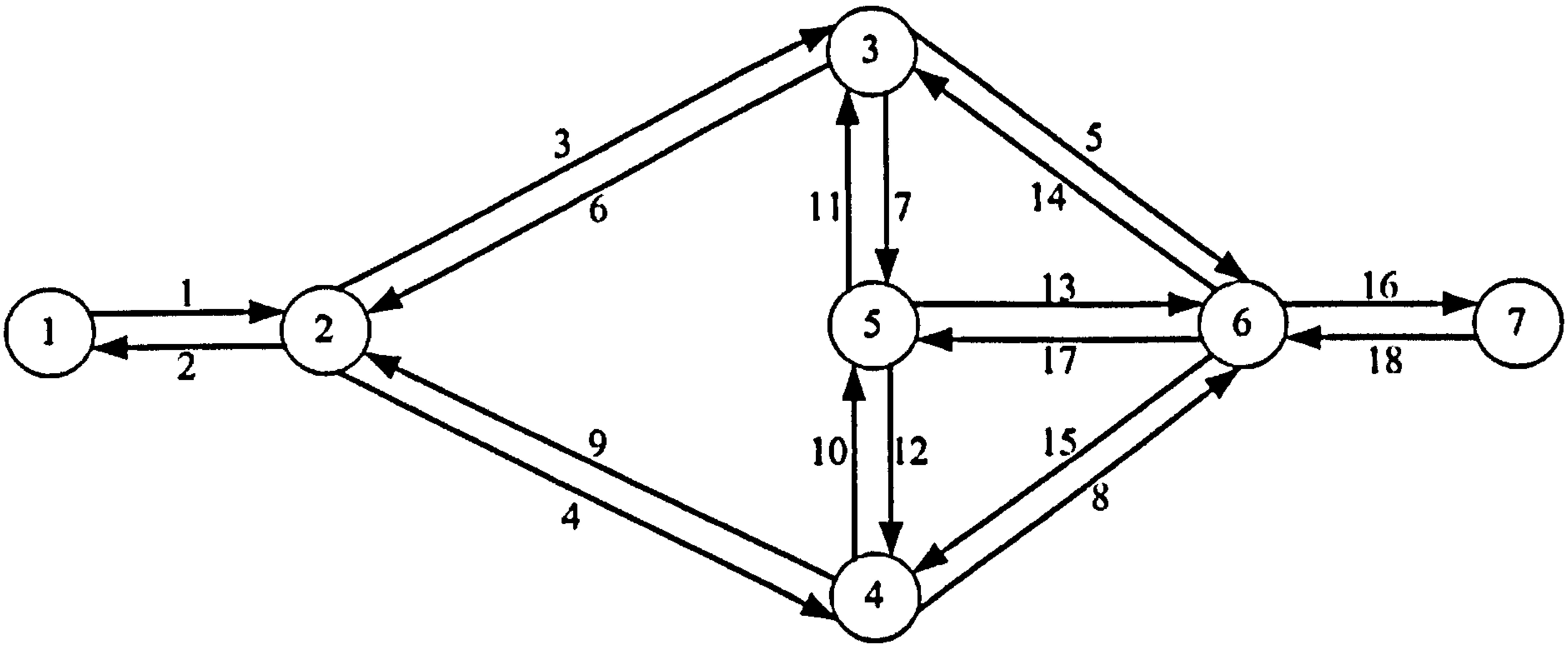


Figure 5-9 Structure of Pacman network for the tests with the improved CPA

Link Number	Starting node	Ending node	a	b	c	n
1	1	2	45.0	9.55	1800	4.5
2	2	1	45.0	9.55	1800	4.5
3	2	3	108.0	108.00	1100	3.0
4	2	4	120.0	120.00	1100	3.1
5	3	6	270.0	57.27	1100	3.5
6	3	2	108.0	108.00	1100	3.0
7	3	5	90.0	90.00	1100	3.2
8	4	6	274.5	58.23	1100	3.0
9	4	2	120.0	120.00	1100	3.1
10	4	5	96.0	96.00	1100	3.1
11	5	3	90.0	90.00	1100	3.2
12	5	4	96.0	96.00	1100	3.1
13	5	6	72.0	72.00	1100	3.1
14	6	3	270.0	57.27	1100	3.5
15	6	4	274.5	58.23	1100	3.0
16	6	7	45.0	9.55	1800	4.5
17	6	5	72.0	72.00	1100	3.1
18	7	6	45.0	9.55	1800	4.5

Table 5-4 Link travel time function parameters for Pacman network

Origin	Destination	d_k^0	μ_k^0
1	5	637	1125
1	7	1027	1050
5	1	522	675
5	7	391	600
7	1	964	1050
7	5	442	850

Table 5-5 Parameters of the demand functions for Pacman network

Test with single toll link

The first set of tests is to optimise the toll level for each link in the network in turn with the objective of maximising the social welfare function (see Section 4.4). From the total number of 18 links, five of them yield no benefit (with the restriction of positive toll) which are link 5, 8, 9, 14 and 15. The improved CPA is applied to optimise the tolls for the remaining 13 links. Table 5-6 below contains the results for the optimal single toll problems.

Link no.	Optimal toll (second)	Social welfare improvement (seconds)	rank
1	495.06	86,280	4
2	164.67	18,215	11
3	141.61	19,443	10
4	106.11	17,011	12
5	-	no benefit	14
6	31.75	5,180	13
7	143.86	95,767	3
8	-	no benefit	14
9	-	no benefit	14
10	131.4	63,762	7
11	107.11	66,556	6
12	93.17	32,108	9
13	179.93	168,463	2
14	-	no benefit	14
15	-	no benefit	14
16	556.00	72,631	5
17	189.27	179,486	1
18	562.75	33,087	8

Table 5-6 Optimal single toll results with the Pacman network

From the table, the improved CPA successfully found the optimal toll level (local optimal at least) for all 13 links. The ranking of the level of the benefit generated by each link is also provided in the table.

Based on Shepherd and Sumalee (2004), the Lagrangian based method (Verhoef, 2002) which relies on the convergence of the UE condition failed to solve the optimal toll problem for link 4 (2-4) due to the convergence error of the equilibrium condition. The improved CPA does not rely on the convergence of the UE condition to determine the improving direction of the toll; hence it is expected that this problem should be resolved with the CPA. Figure 5-10 shows the iterations of the improved CPA and the curve of the objective function as the function of the toll on link 4 (link 2-4).

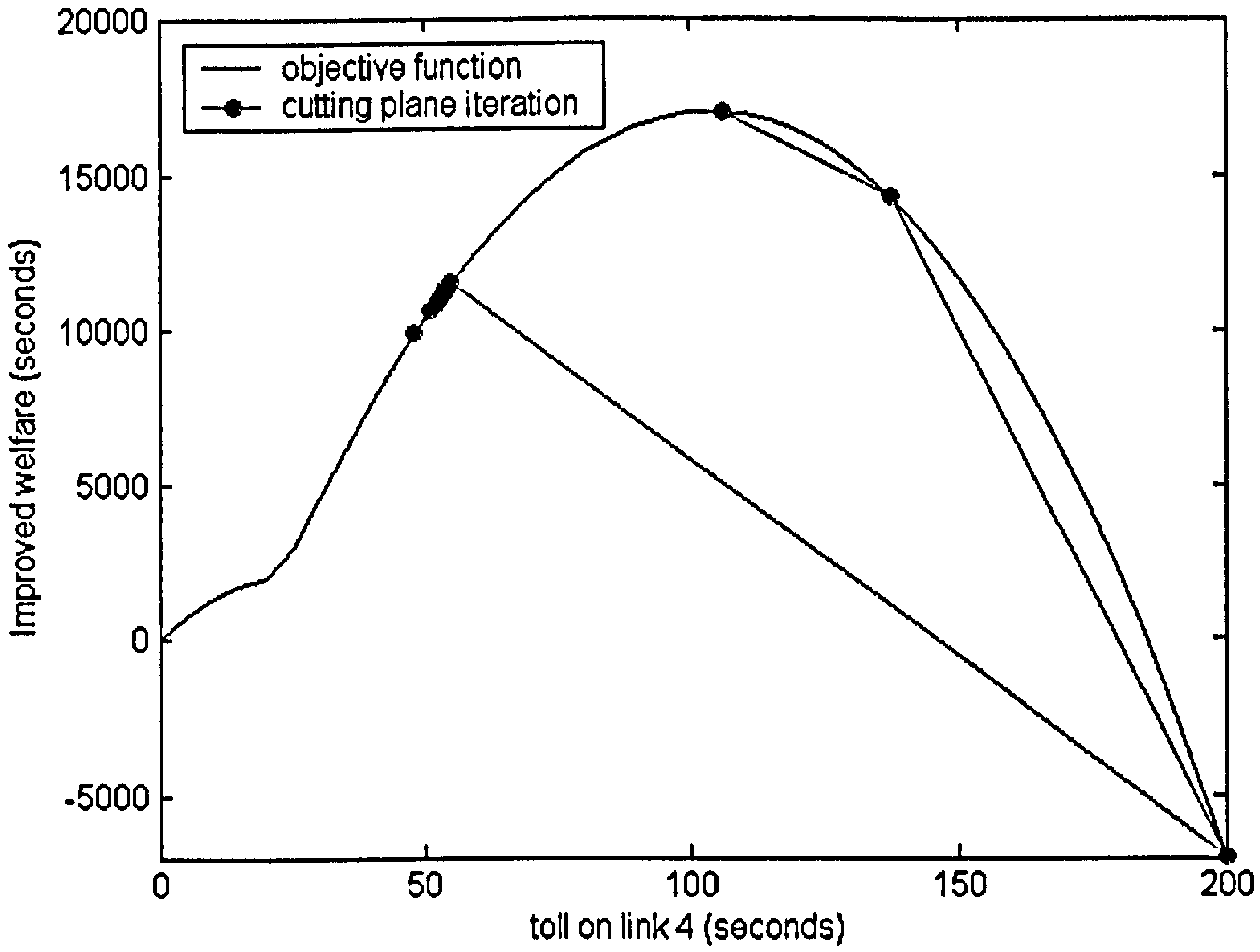


Figure 5-10 Iteration of the improved CPA method optimising the toll on link 2-4 (Pacman)

As reported in Shepherd and Sumalee (2004), the Lagrangian based algorithm converged to the optimal toll of around 27 seconds which is neither a local optimum nor stationary point. On the contrary, the improved CPA managed to find the optimal toll level of around 106 seconds which can be certified by Figure 5-10. Figure 5-11 shows the iterations of the CPA optimising the toll on link 18 (link 7-6), which produced the highest net social welfare improvement, and the curve of the objective function.

From Figure 5-11, there exist several local optima. The improved CPA incidentally found the global optimal toll level of around 560 seconds. From the numerical experience, the starting point of the algorithm may help the improved CPA to find the global optimum rather than becoming stuck at a local optimum. A strategy adopted in the tests is to introduce the initial point as the system optimum solution which can be solved easily.

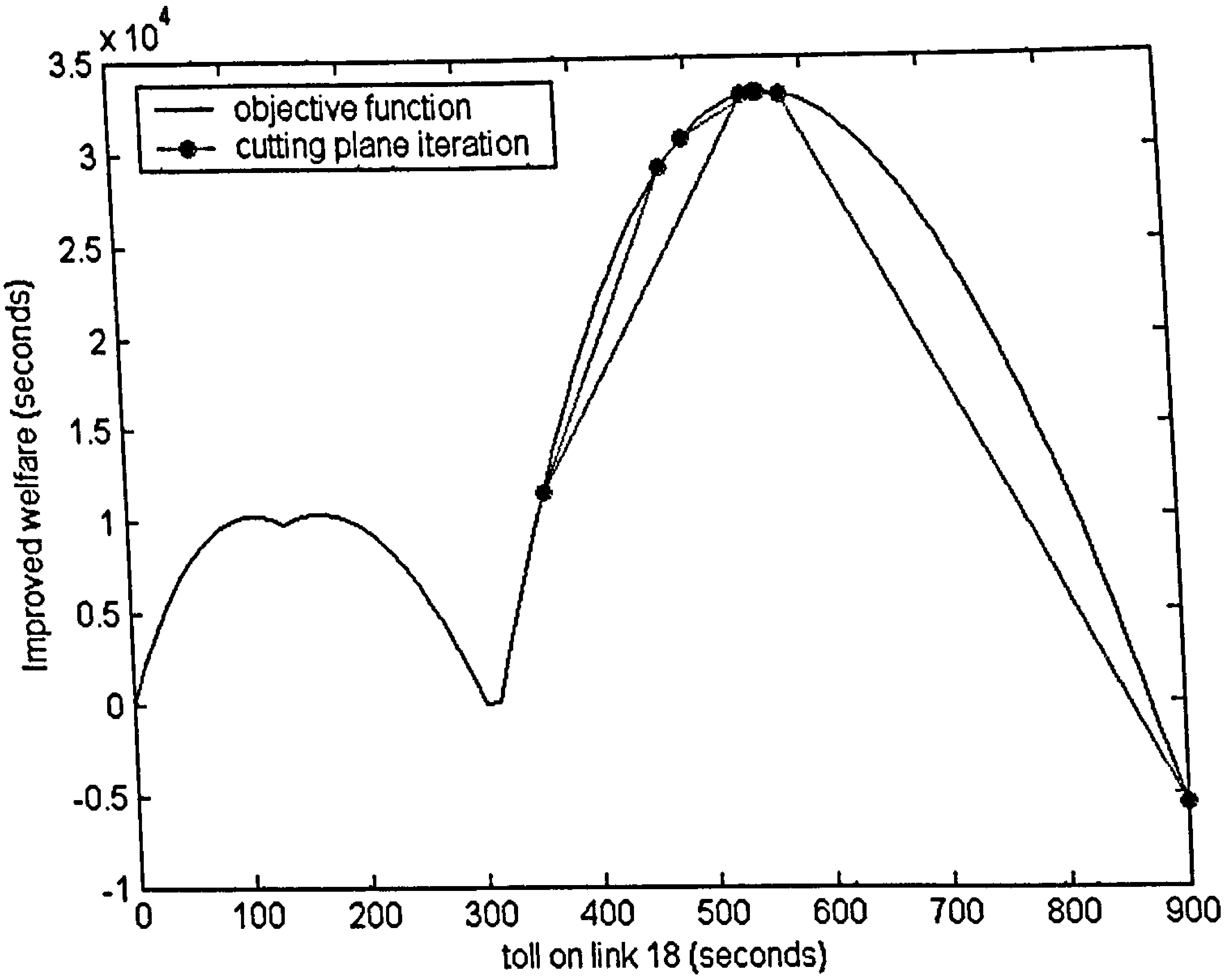
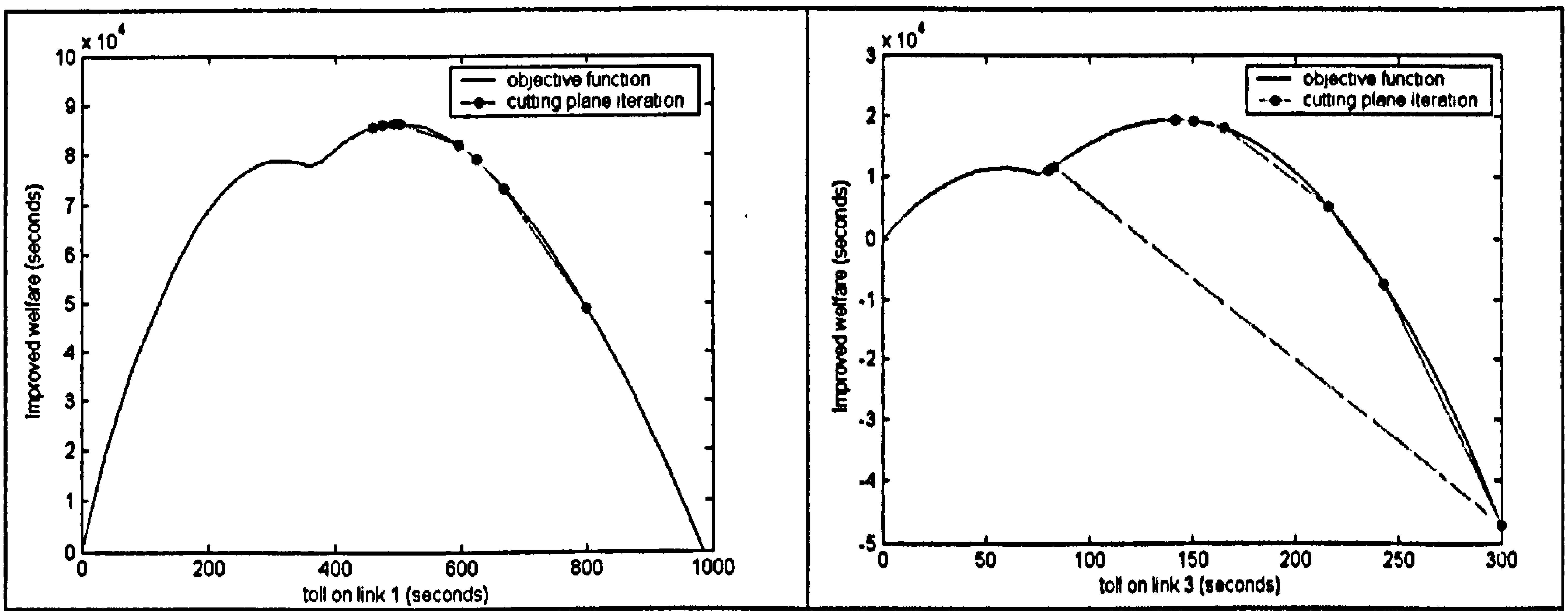


Figure 5-11Iteration of the improved CPA method optimising the toll on link 7-6

Figure 5-12 below shows the results from the tests with other tolled links: link 1, link 3, link 4, and link 13. The figure shows that in all tests the improved CPA converged to the true optimal solutions of those problems.

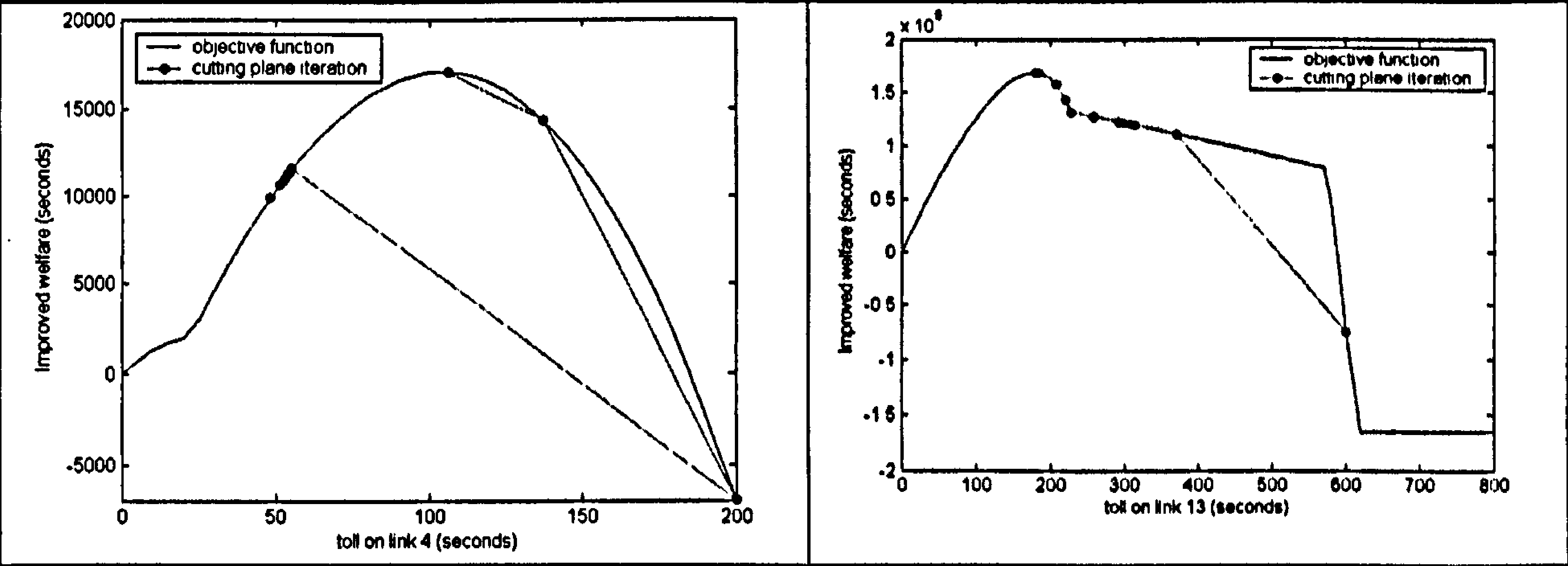
Interestingly, from the numerical experiences when the improved CPA is applied to the relaxed optimal toll problem (with a low number of extreme points) the solution in terms of link flow, demand flow, and tolls converged to the system optimum initially before finally reaching the real optimal toll (which is often the global optimum as well).

By relaxation we mean not all the extreme points are included in the VI constraint, and hence there is less restriction on the UE condition. One conjecture could be made that with the relaxed optimal toll problem as defined in Equation 5-9, in the early iteration the problem in Equation 5-9 becomes the system optimal problem which provides the upper bound of the original COTP problem. Hence, with this possibility the initial solution of Equation 5-9 often becomes the system optimal flow and gradually the additional constraints on the VI force the algorithm to converge to the true optimal solution of the COTP within the vicinity of the system optimal solution. However, it should be stressed that there is no concrete proof for this conjecture at this stage and this strategy should be treated purely as a heuristic approach.



toll on link 1

toll on link 3



toll on link 4

toll on link 13

Figure 5-12 Examples of the iterations of the improved CPA (link 1, 3, 4, and 13)

Test with two and three tolled links

Now, the improved CPA is tested with a problem with two tolled links. Table 5-7 below shows the optimisation result with the two and three tolled links. The first three pairs are chosen from their ranking of their individual tolled link and the three tolled links in the last test forming a charging cordon around the city centre (node 5).

Figure 5-13 and Figure 5-14 show examples of the objective function contour and surface and the optimisation iterations for test number 1 and 3). As illustrated in the figures, in both cases there exist a number of local optima. The plots of the improved CPA iterations show the convergence of the algorithm to the actual optimal solution.

Test no.	Tolled links	Optimal tolls (seconds)	Social welfare improvement (seconds)
1	17,13	184.21, 178.31	348,218
2	17,7	184.35, 141.96	275,535
3	17,1	184.28, 466.45	265,988
4	7,10,17	173.35, 170.98, 184.37	381,325

Table 5-7 Test results with two and three tolled links (Pacman network)

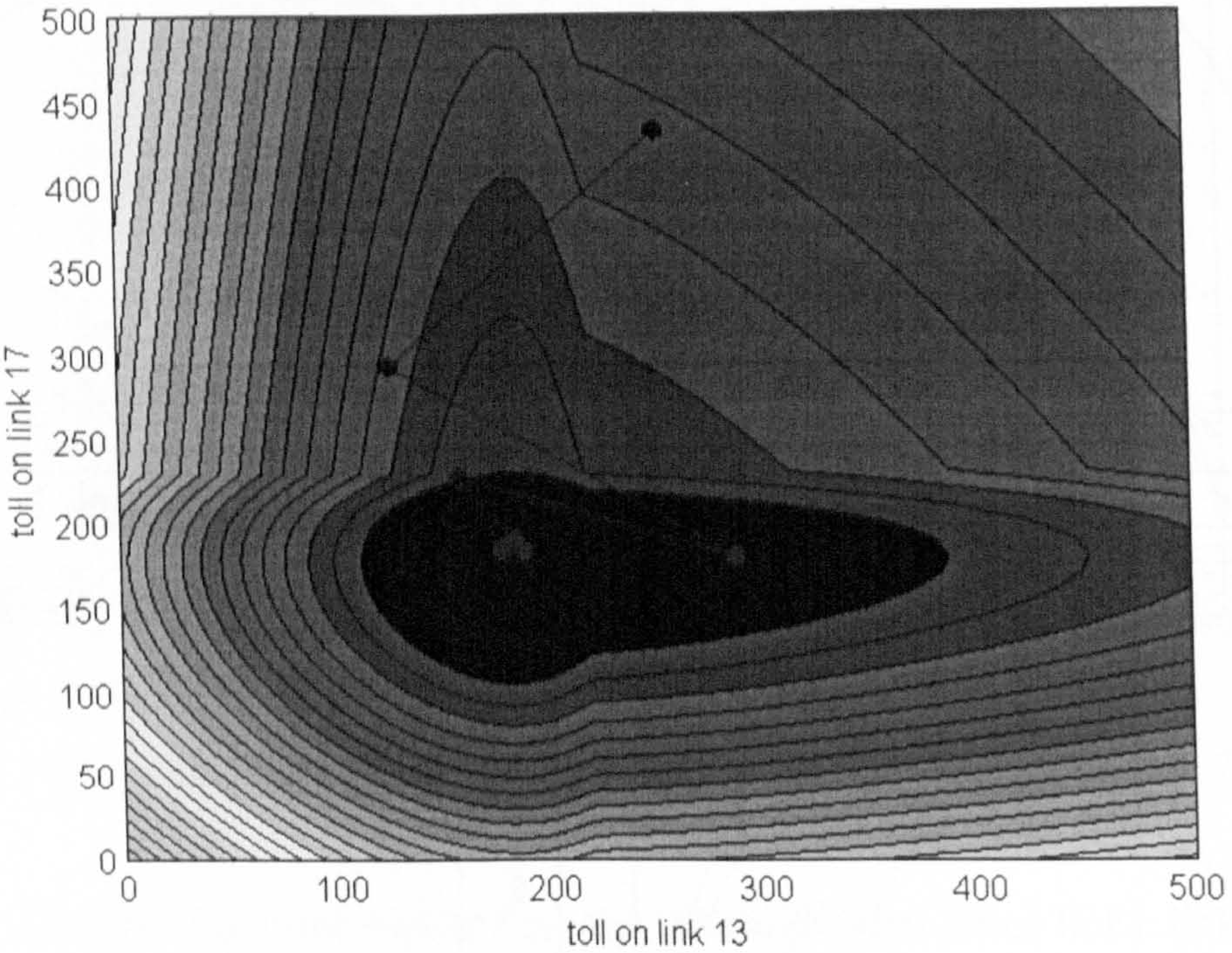


Figure 5-13 Contour plot of the objective function and iterations of the improved CPA (link 17 and link 13)

To illustrate the concept of the relaxed equilibrium condition, Figure 5-15 shows the convergence of the objective function from the Master Problem (with relaxed equilibrium) and the objective function recalculated at the true UE condition; and Table 5-8 contains the value of the gap function from each iteration.

As expected, the Master Problem objective function gradually converged to the true objective function as the number of extreme points added into the *VI* increases. The objective function from the Master Problem can be lower than the true objective function value in the early iterations of the improved CPA. Given the relaxed UE condition, the solution from the Master Problem should be the upper bound of the optimal toll problem. In the extreme case with no *VI* constraint added in the Master Problem the solution should be the system optimal flows hence giving the upper bound of the optimal toll problem. However, this may not be the case for the improved CPA since the relaxed feasible region is also introduced in the feasible flow space. In this case, the possible set of flows may be too limited to yield the system optimal flow. On the other hand, the original CPA will give the upper bound for the optimal toll problem since the feasible flow space is represented fully.

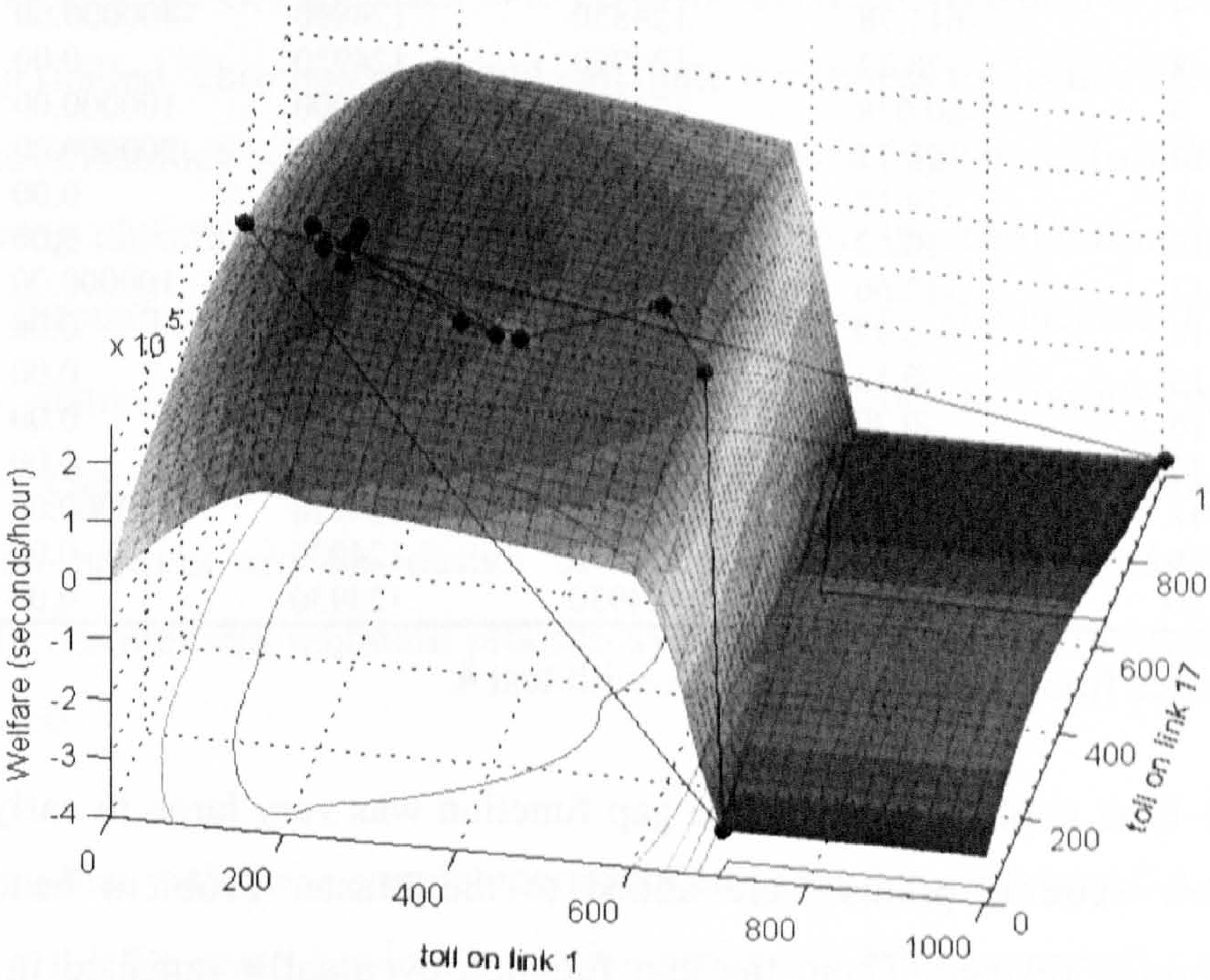


Figure 5-14 Objective function surface and iterations of the improved CPA (link 17 and link 1)

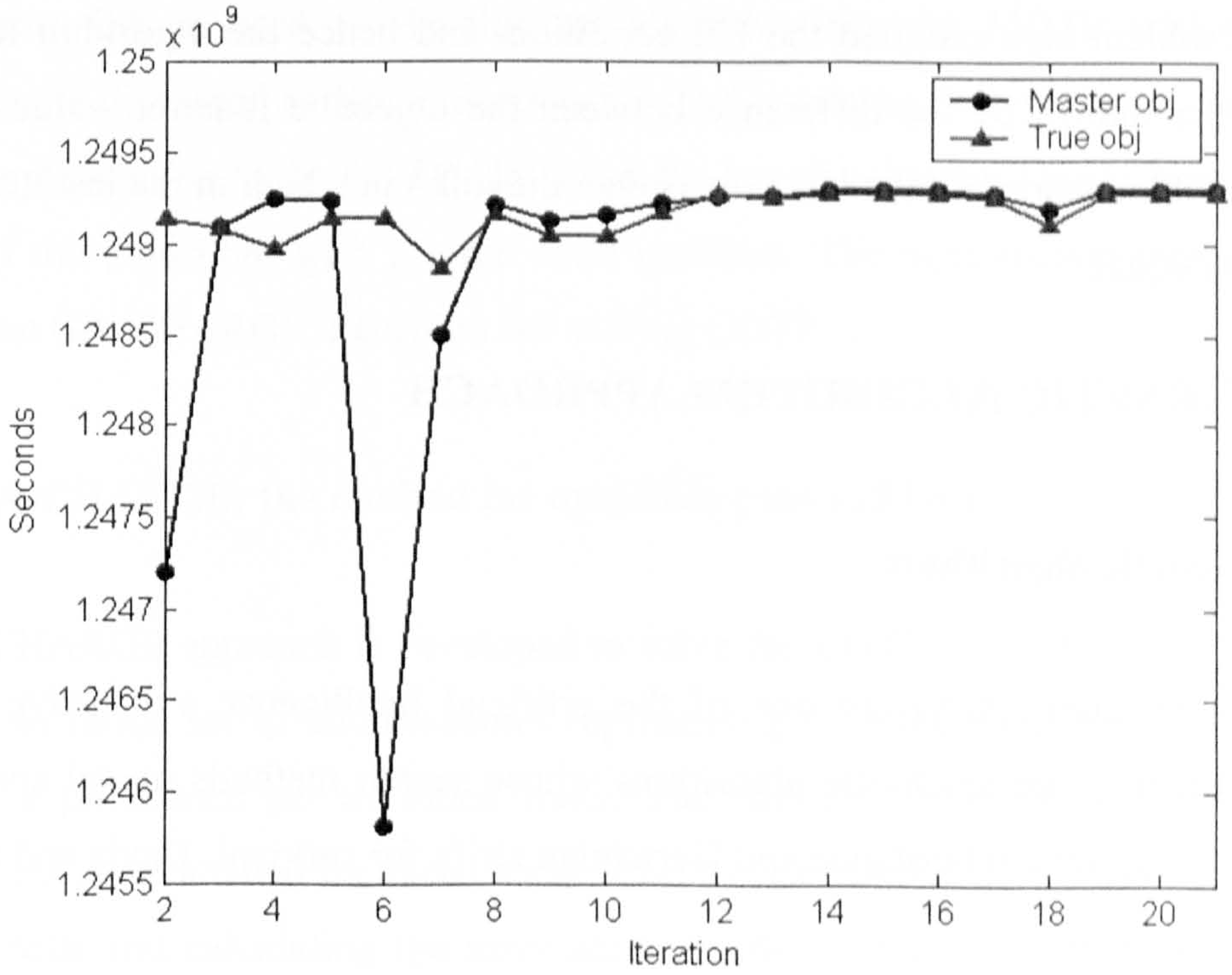


Figure 5-15 Convergence of the Master objective and True objective functions with CPA

Iter.	Gap function (10^4)	Master objective(10^4)	UE objective(10^4)	col 4 – col 3
1	-10020000.00	104830	124930	-201000000.00
2	-407.75	124720	124910	-1900000.00
3	-182.00	124910	124910	0.00
4	-82.778	124930	124900	300000.00
5	-37.83	124920	124910	100000.00
6	-835.13	124580	124910	-3300000.00
7	-611.78	124850	124890	-400000.00
8	-78.27	124920	124920	0.00
9	-80.038	124910	124900	100000.00
10	-70.77	124920	124900	200000.00
11	-24.15	124920	124920	0.00
12	-10.62	124930	124930	0.00
13	-13.66	124930	124920	100000.00
14	-0.99	124930	124930	0.00
15	-0.13	124930	124930	0.00
16	-0.20	124930	124930	0.00
17	-0.097	124930	124930	0.00
18	-98.01	124920	124910	100000.00
19	0.00	124930	124930	0.00
20	0.00	124930	124930	0.00

Table 5-8 Iterates from the improved CPA with test 4

From Table 5-8, as explained earlier the gap function was very large in early iterations since only few extreme points were added to the Master Problem hence the UE condition is highly relaxed. Then, the gap function eventually vanished in which the new shortest path or extreme flow was not the descent direction for the UE problem (the objective of the Sub Problem is greater than or equal to zero). Hence, the solution to the Master Problem also satisfied the UE condition, and hence the algorithm terminated. This is also shown by the difference between the objective function values from the Master Problem and with the UE flow (given the tolls) in which in the last iteration two values are equal.

5.6 GENETIC ALGORITHM APPROACH

5.6.1 Genetic algorithms

Genetic algorithms (GA) are one of the artificial intelligence exhaustive searching techniques; they are stochastic algorithms whose search methods model some natural phenomena: genetic inheritance and Darwinian strife for survival. Davis and Steenstrup (1987) stated that:

"The metaphor underlying genetic algorithms is that of natural evolution. In evolution, the problem each species faces is one of searching for beneficial adaptations to a complicated and changing environment .The 'knowledge' that each species has gained is embodied in the makeup of the chromosome of its members."

The basic idea of the GA approach is to code the decision variables of the problem as a finite string (called 'chromosome') and calculate the fitness (objective function) of each string. Chromosomes with a high fitness level have a higher probability of survival. The surviving chromosomes then reproduce and form the chromosomes for the next generation through the 'crossover' and 'mutation' process. The method of GA is widely applied in many disciplines, but most applications have to modify the GA to the problem or change the problem to be compatible with GA. The main parts in the modification process are the design of chromosome encoding and of the genetic operators (crossover and mutation processes) in order to maintain the search within the feasible space.

It is amazing that with these simple operations and searching strategy GA can produce a very good result for many hard optimisation problems. Many successful applications of GA in different disciplines have been reported. This is the key stimulus for the development of the GA based (GAB) optimisation algorithms for the COTP (also in the next chapter GA is used to develop an algorithm for the MOTP with topology constraint). As discussed earlier in Section 5.2, this is not the first attempt to implement the GAB optimisation to the COTP. However, the key development made here is indeed the test of the algorithm with a large scale problem. The next section introduces the detail of the GA-CHARGE algorithm for solving COTP.

5.6.2 GA-CHARGE: the method for optimising the toll level

The GA-CHARGE approach is developed to solve the COTP. GA-CHARGE randomly generates an initial set of chromosomes representing possible combinations of charge levels on a predefined set of tolled links. The benefits in terms of social welfare improvement are evaluated for each charge level by solving the UE assignment problem given the tolls and calculating the associated benefits. GA-CHARGE then selects the parent chromosomes for the next generation based on the performance of each chromosome. Since the fitness value in GA-CHARGE can be negative, the selection is based on the linear ranking roulette wheel selection process (Michalewicz, 1992). The detail of this selection process is provided in Section 6.4.4 (page 185) in Chapter 6.

In addition, the elitism preservation strategy is also used in which the chromosomes in the current generation are ranked according to their fitness, and then the best *n* chromosomes are automatically passed on to the next generation without any distortion from the crossover and mutation operators (the user needs to specify how many chromosomes he/she want to preserve). Nevertheless, the elitism chromosomes are still maintained in the set of chromosomes for the normal operations of selection, crossover, and mutation process to allow them to exchange their good genes to other chromosomes. The genetic operators, crossover and mutation, are then randomly applied to the parents to produce the offspring for the next generation. Figure 5-16 shows the overall framework of GA-CHARGE. Since the locations of tolled links are predefined in advance, the objective function does not need to include the operator's cost of the charging scheme. Next the overall optimisation process of GA-CHARGE, chromosome encoding, and genetic operators (crossover and mutation) are explained.

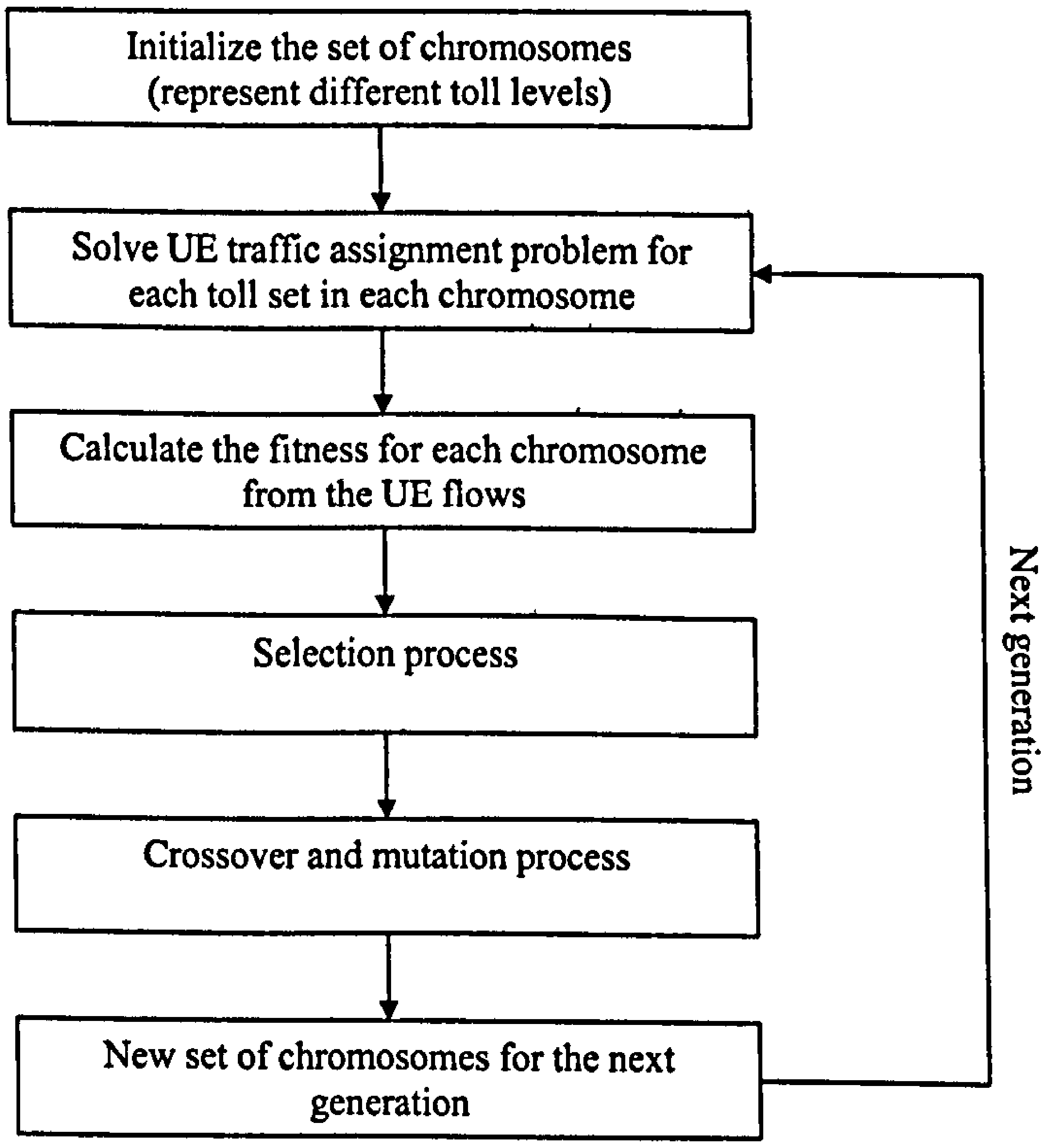


Figure 5-16 Overall framework of GA-CHARGE

Chromosome encoding

Let t be the number of predefined tolled links and let r be the predefined maximum toll level. Each chromosome represents a set of toll levels for the t tolled links in binary format. The structure of the chromosome is therefore a matrix A with t columns and k rows where k is determined by the number of digits required to represent the maximum toll in binary format. Figure 5-17 shows an example chromosome (A matrix) for ten tolled links. The toll on each link is defined by the binary number in each column which is shown in the bottom row.

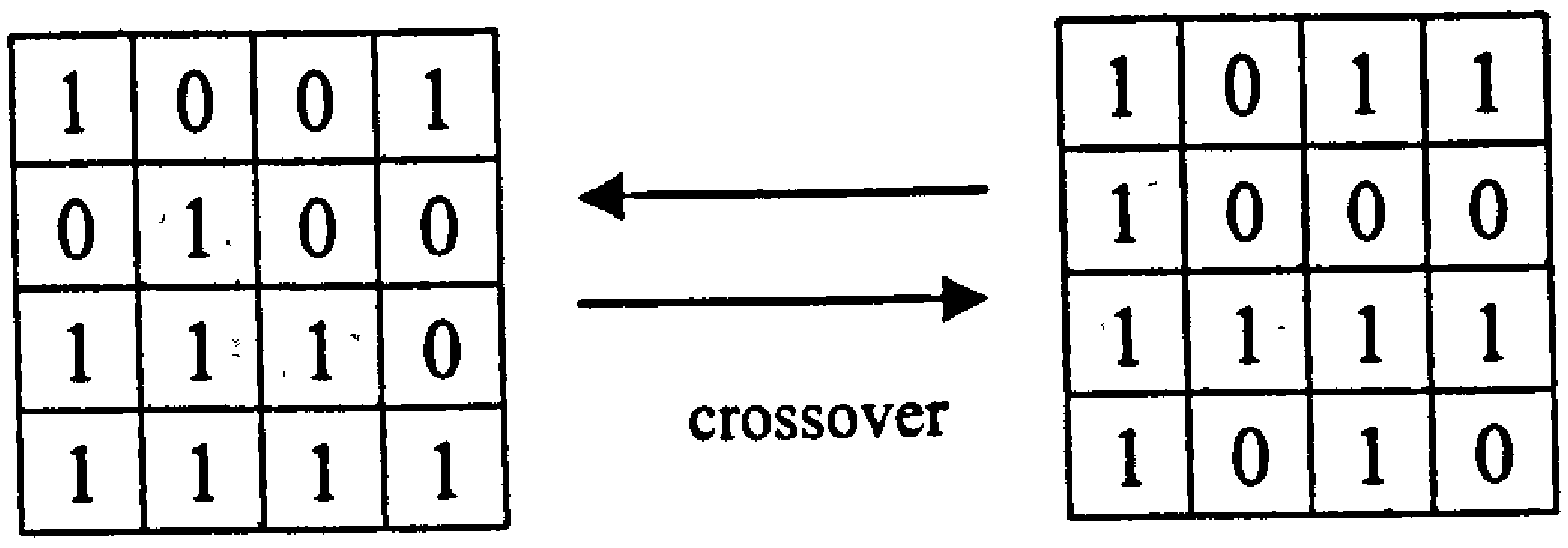
$$\begin{array}{rc} A = & \left[\begin{array}{cccccccccc} 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \end{array}\right] \\ & \begin{array}{cccccccccc} 31 & 21 & 3 & 16 & 11 & 20 & 15 & 10 & 19 & 22 \end{array} \end{array}$$

Figure 5-17 Chromosome structure for GA-CHARGE

The equivalent number of the binary number in each column of A represents the level of the toll on the corresponding tolled link of that column. In order to reduce the complexity of the problem, the possible toll levels are pre-specified in advance. For instance, the maximum of 31 possible discrete toll levels can be specified for the matrix A in Figure 5-17 (e.g. the toll level from 0-3100 pence with 100 pence interval between each toll level). With the example in Figure 5-17, the tolls from this matrix are 3100, 2100, 300, 1600, 1100, 2000, 1500, 1900, and 2200 in that order.

Crossover and mutation process

At the beginning of the crossover process, the chromosomes in the current generation are randomly mated before the crossover process is carried out. Figure 5-18 shows the crossover process in GA-CHARGE. The crossover process is to select at random a partition from the chromosome matrix (highlighted area) which is then switched between two "mated" chromosomes.



1	0	0	1
0	1	0	0
1	1	1	0
1	1	1	1

1	0	1	1
1	0	0	0
1	1	1	1
1	0	1	0

Figure 5-18 Crossover process in GA-CHARGE

After the crossover process, the mutation process is applied to the offspring. The mutation process randomly chooses cells to be "mutated". If selected, the value in that cell is changed from 0 to 1 or vice-versa.

There are two main parameters the user needs to define namely the Probability of crossover (Pc) and Probability of mutation (Pm). For the crossover, for each mated chromosome a random number is generated and if the random number is less than the given Pc, then the crossover process is carried out. Similarly, for each element of the chromosome (A) a random number is generated and if the random number is less than the given Pm, the value in that element is changed to 1 if the current value is 0 and vice versa.

5.6.3 Experiments with GA-CHARGE

Firstly, GA-CHARGE is tested with the Pacman network (see Figure 5-9). The tests with two and three tolled links with the Pacman network with the CPA are reproduced here but with GA-CHARGE. The parameters for GA-CHARGE are set as follows:

Case	population size	generation number	Pc	Pm
2 links	20	5	0.60	0.05
3 links	50	50	0.60	0.05

Table 5-9 Parameters used for GA-CHARGE with Pacman test

Note that the toll interval of 10 seconds is adopted with 63 possible steps (hence the possible toll levels range from 0 – 630 seconds). Six binary elements representing the toll level were allowed for each tolled link. Thus, the total number of possible combinations of the toll levels for the two and three link cases is 3,969 and 250,047 respectively. Therefore, the number of chromosomes tested in the two and three link cases as shown in Table 5-9 is approximately 2.5% and 1% of the total number of possible combinations in that order.

Table 5-10 presents the test results with the Pacman network using GA-CHARGE. The results show that GA-CHARGE is capable of solving the COTP. In all tests, GA-CHARGE could identify the solution near to the solutions found by the CPA with some discrepancy that may be due to the discretised toll levels. The CPU times spent by GA-CHARGE in each of these four tests are slightly longer than those tests with CPA as expected.

Test no.	Tolled links	Optimal tolls (seconds)		Objective function (seconds)	
		CPA	GA-CHARGE	CPA	GA-CHARGE
1	17,13	184.21, 178.31	170.00, 170.00	348,218	345,239
2	17,7	184.35, 141.96	190.00, 140.00	275,535	275,243
3	17,1	184.28, 466.45	190.00, 440.00	265,988	256,711
4	7,10,17	173.35, 170.98, 184.37	170.00, 160.00, 180.00	381,325	379,520

Table 5-10 GA-CHARGE results with Pacman network

Figure 5-19 plots the chromosomes from the first test labelled by the generation number. This figure demonstrates the evolution process in GA-CHARGE. In early generations, the chromosomes were spread over a wide space of the objective function surface. Then, with the genetic operators and natural selection process GA-CHARGE gradually produced chromosomes with fitter and fitter objective function. Eventually (in this case in the fifth generation), GA-CHARGE found the optimal solution of the problem.

Figure 5-20 below shows a different perspective of the process of GA-CHARGE using the result from test no. 4. The figure shows the plots of the best solution found so far and the average fitness (objective function) in each generation.

Next, GA-CHARGE is tested with the Edinburgh network. Figure 5-21 shows the network of the Edinburgh city in U.K. The network has totally 344 links and 525 OD pairs.

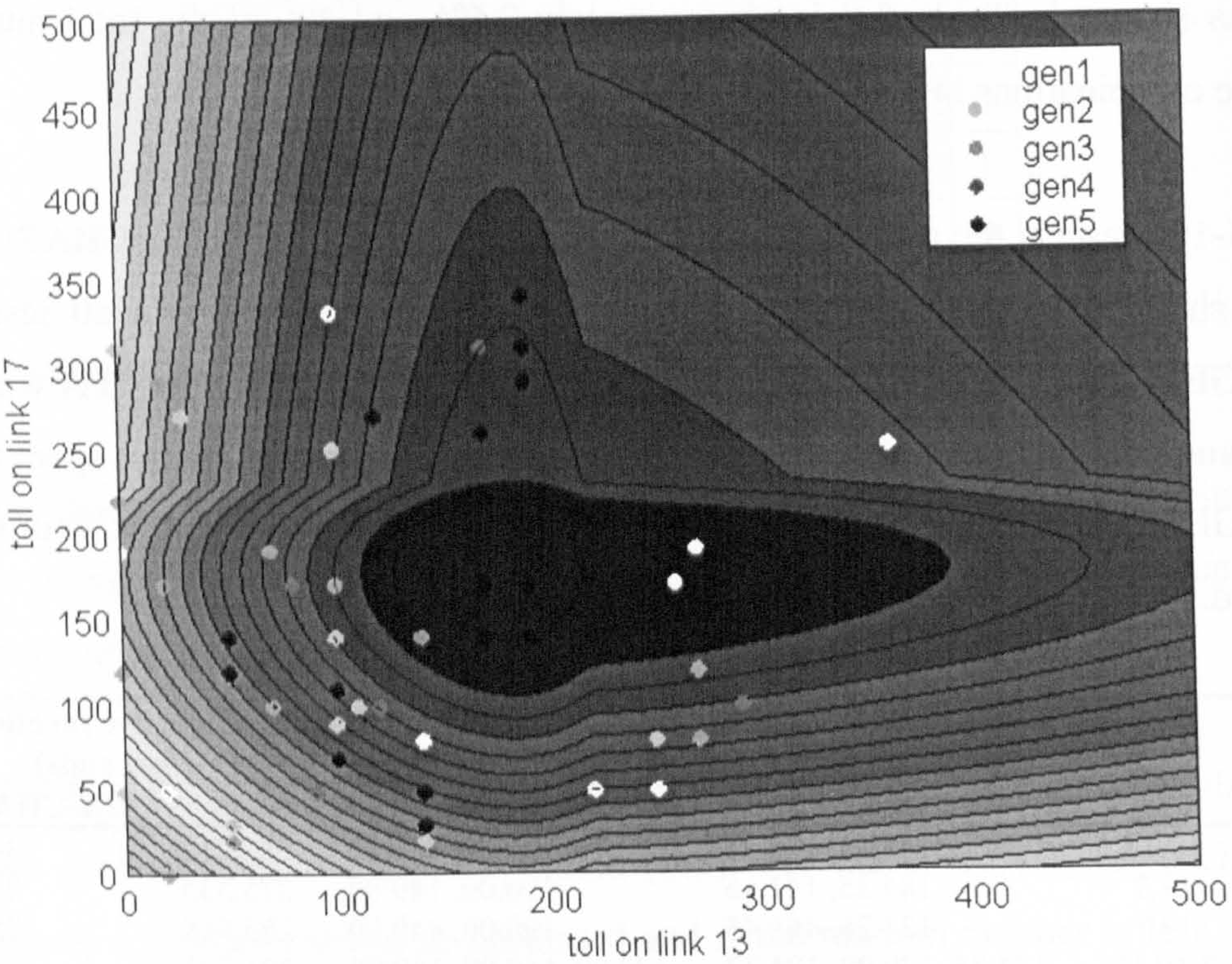


Figure 5-19 Chromosomes generated by GA-CHARGE for test no. 1 with Pacman network

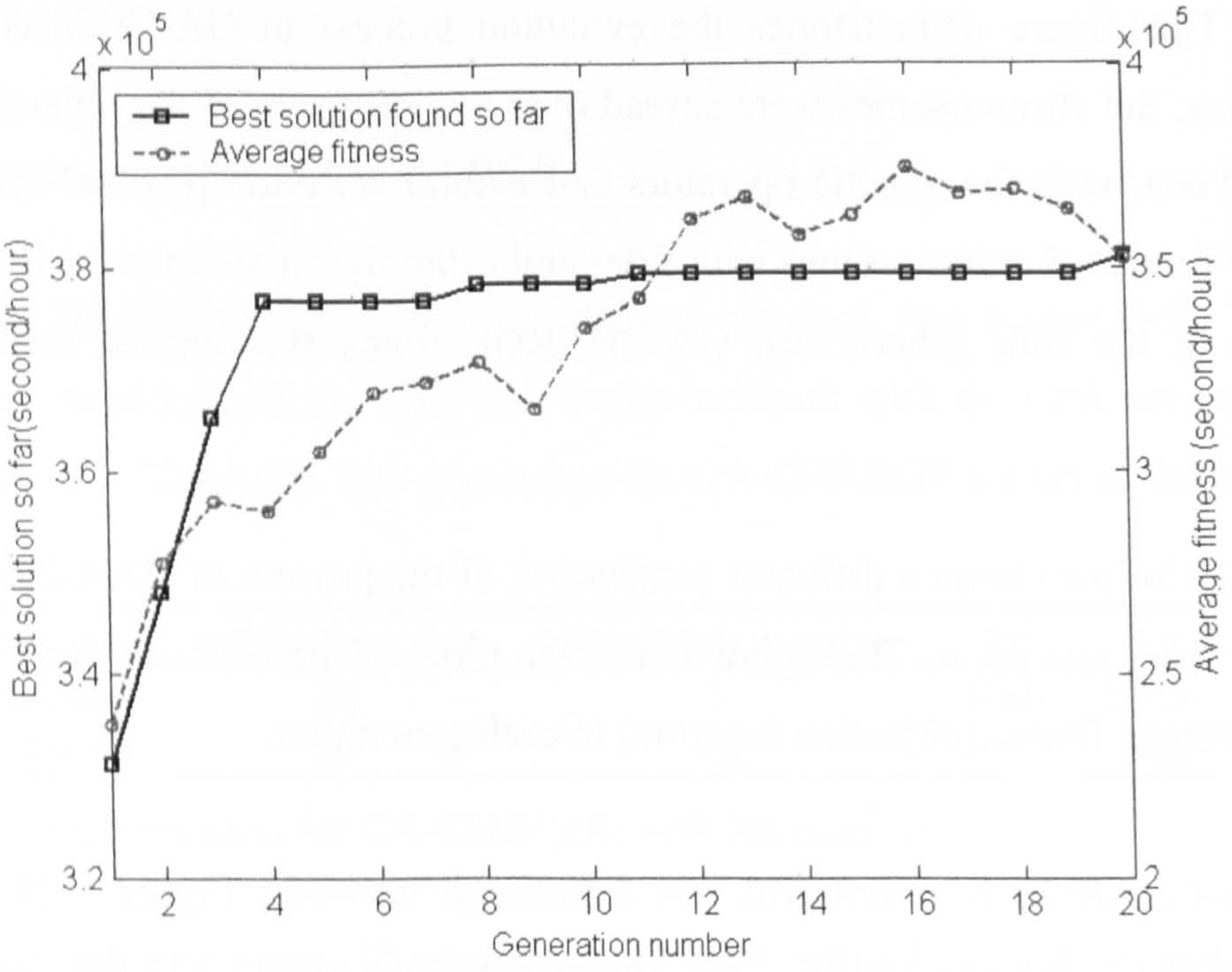


Figure 5-20 Best solution found so far and average fitness in each generation from the test of GA-CHARGE with Pacman network (test no. 4)

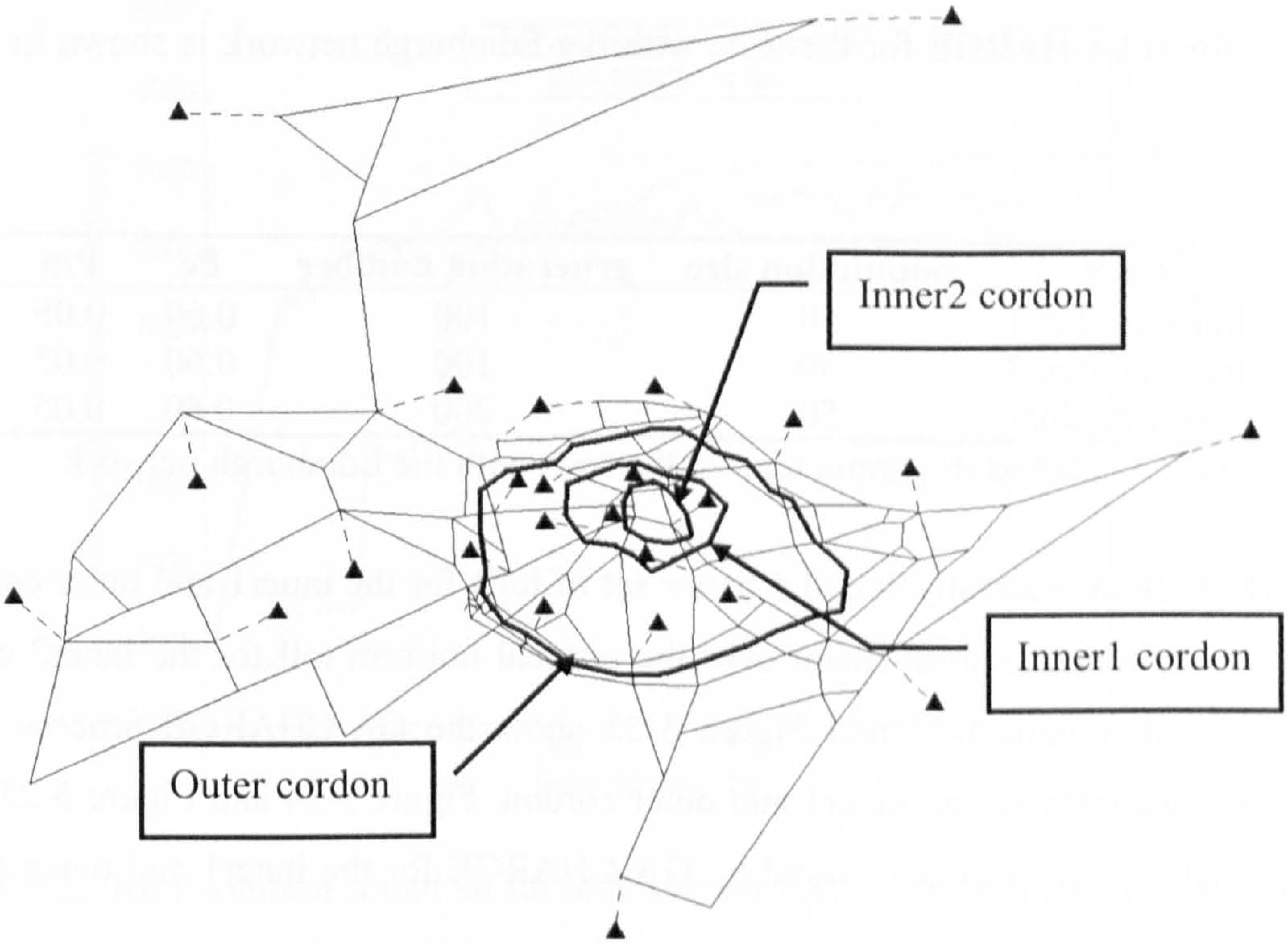


Figure 5-21 Edinburgh network with three judgmental cordons

The link travel cost is the same function adopted in the Pacman network:

$$t_j = a + b \cdot \left(\frac{v_j}{c} \right)^n .$$

Similarly, the OD demand function is:

$$d_k = d_k^0 \cdot \left(\frac{\mu_k}{\mu_k^0} \right)^\beta .$$

The parameters for both link cost and OD demand function for the Edinburgh network are shown in Appendix C. The value of time and vehicle running cost rate are 7.63 pence per minute and 5.27 pence per kilometre. These parameters are recommended in Department of Transport(1989). For comparison with the improvement of the variable toll scheme, a uniform toll for each cordon is optimised as well as a variable toll around the cordon (using GA-CHARGE). The optimal uniform tolls for the inner1, inner2, and outer cordons are £0.50 (393 seconds), £0.75 (589 seconds), and £0.75 (589 seconds) respectively. The capital and operating costs per toll point were estimated to be £183,400 per charge point and £85,300 per charge point per annum respectively at 2000 prices (Oscar Faber, 2001). This is equivalent to £100 per toll point per peak-hour if charges are assumed to apply over 1000 peak hours per year, the schemes are assumed to have a life of 30 years, and a discount rate of 6 per cent is used. GA-CHARGE is used to optimise the tolls around these three judgmental cordons. The parameters adopted for GA-CHARGE for the tests with the Edinburgh network is shown in Table 5-11.

Case	population size	generation number	Pc	Pm
inner cordon 1	50	100	0.60	0.05
inner cordon 2	50	100	0.60	0.05
outer cordon	50	200	0.60	0.05

Table 5-11 GA-CHARGE parameters for the tests with the Edinburgh network

GA-CHARGE successfully found a better set of tolls for the inner1 and outer cordons. However, no further improvement over the optimal uniform toll for the inner2 cordon can be found. Figure 5-22 and Figure 5-23 show the GA-CHARGE process whilst optimising the tolls for the inner1 and outer cordon. Figure 5-24 and Figure 5-25 show the optimal tolls (in seconds) found by GA-CHARGE for the inner1 and outer cordon respectively.

Table 5-12 contains the results for the tests of GA-CHARGE with the Edinburgh network. As mentioned, the optimal uniform tolls for the inner1 and outer cordons are 393 and 589 seconds in that order. The solutions shown in Figure 5-24 and Figure 5-25 and the comparison of the benefit in Table 5-12 demonstrate the benefit of the variation of the toll levels around the charging cordon. The net benefits for the inner1 and outer cordon with the variable tolls increase by 50% and 74% compared to the uniform toll schemes respectively. However, the variable toll may not outperform the uniform toll in some case, e.g. the case of the inner2 cordon for the Edinburgh network where no additional benefit is found from introducing the variable tolls. This situation is varied and controlled by the condition of the network and the location of the cordon.

Charging cordon	Toll regime	No. of toll points	Cost (£k/hour)	Total benefit (£k/hour)	Net benefit (£k/hour)
Inner cordon 1	uniform	9	0.90	3.00	2.10
	varied	9	0.90	4.07	3.17
Inner cordon 2	uniform	7	0.70	4.69	3.99
	varied	7	0.70	-	-
Outer cordon	uniform	20	2.00	3.96	1.96
	varied	20	2.00	5.41	3.41

Table 5-12 Results from the tests of GA-CHARGE with the Edinburgh network

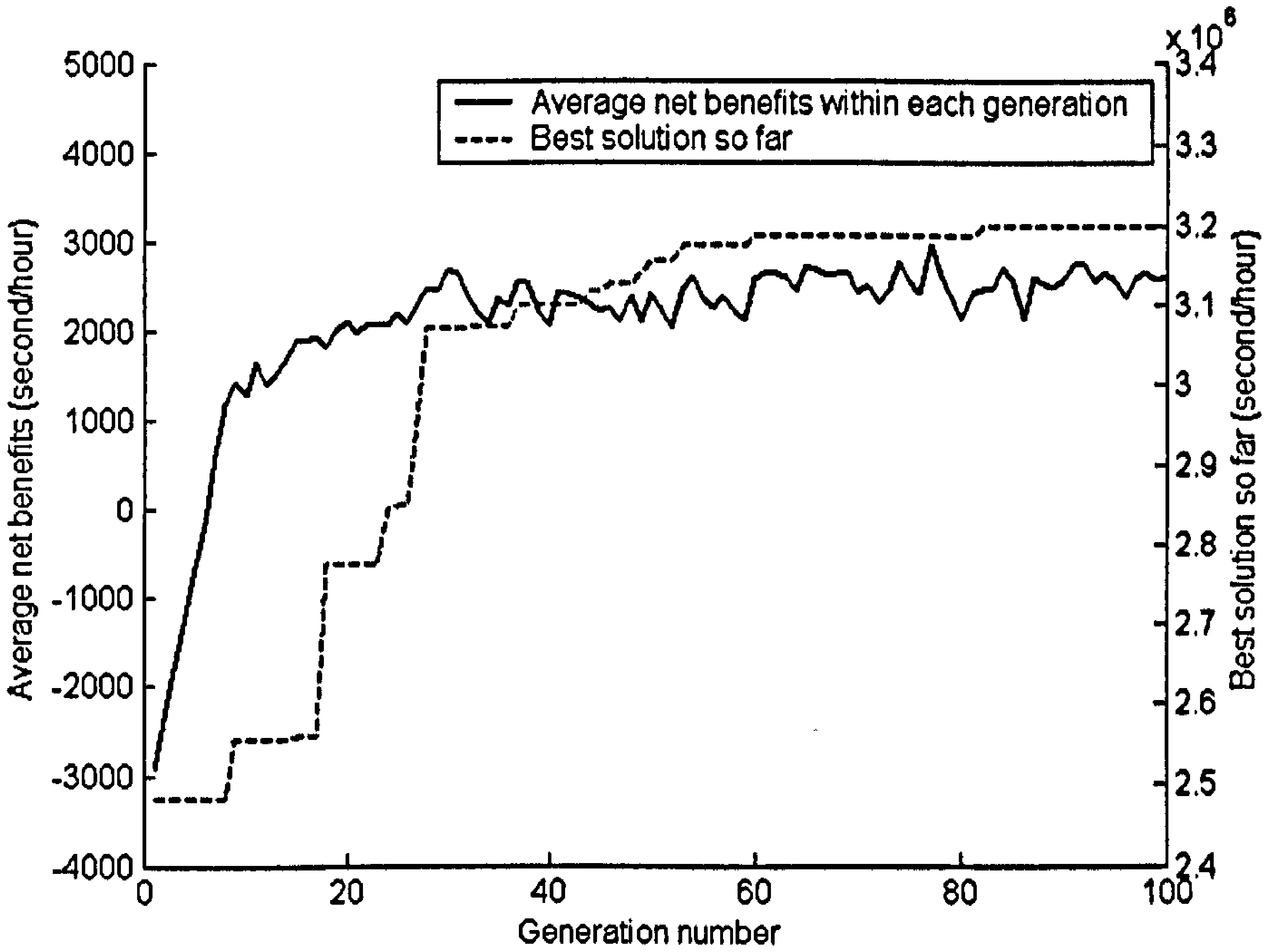


Figure 5-22 Best solution found so far and average fitness in each generation from the test of GA-CHARGE with the Edinburgh network (inner1 cordon)

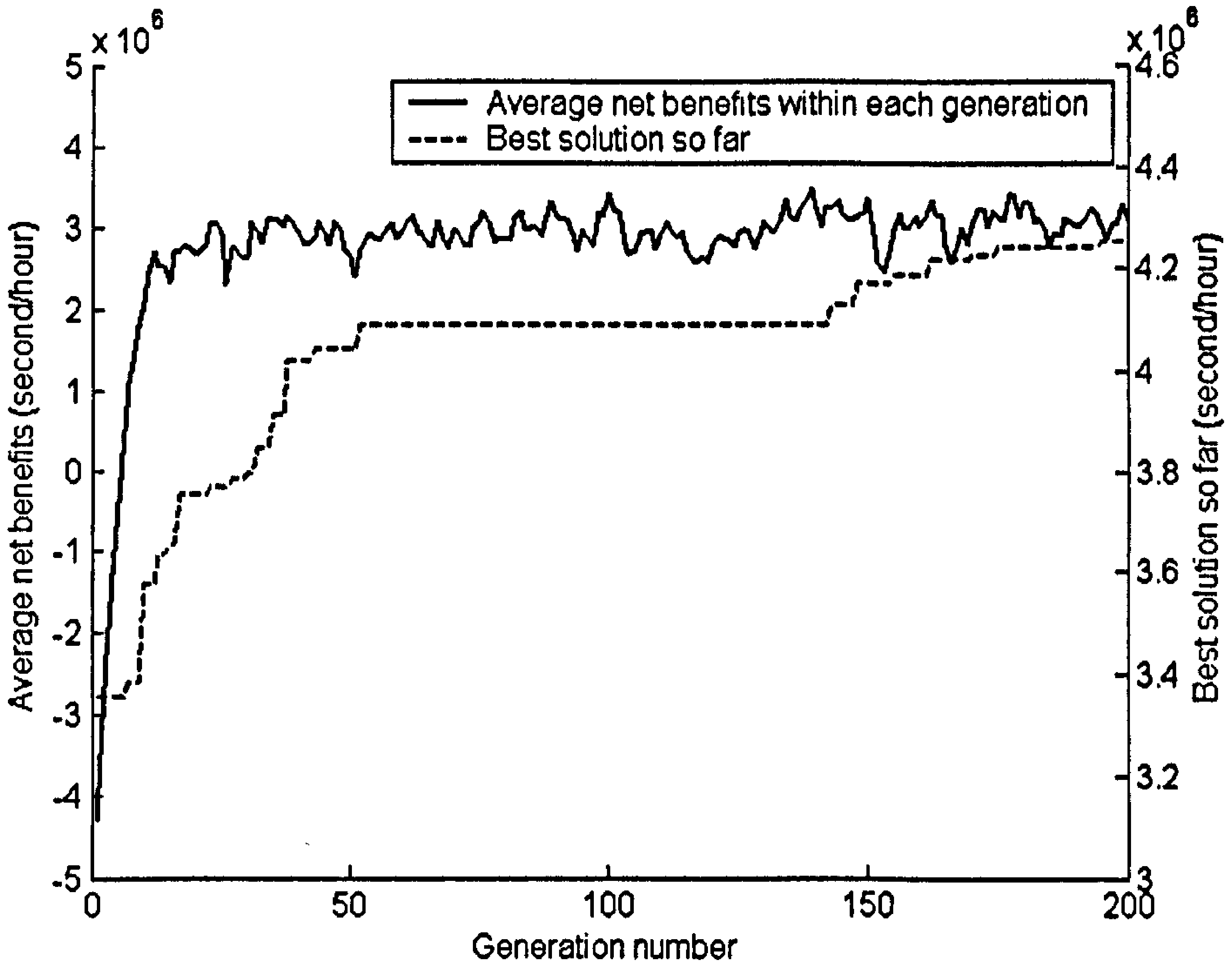


Figure 5-23 Best solution found so far and average fitness in each generation from the test of GA-CHARGE with the Edinburgh network (outer cordon)

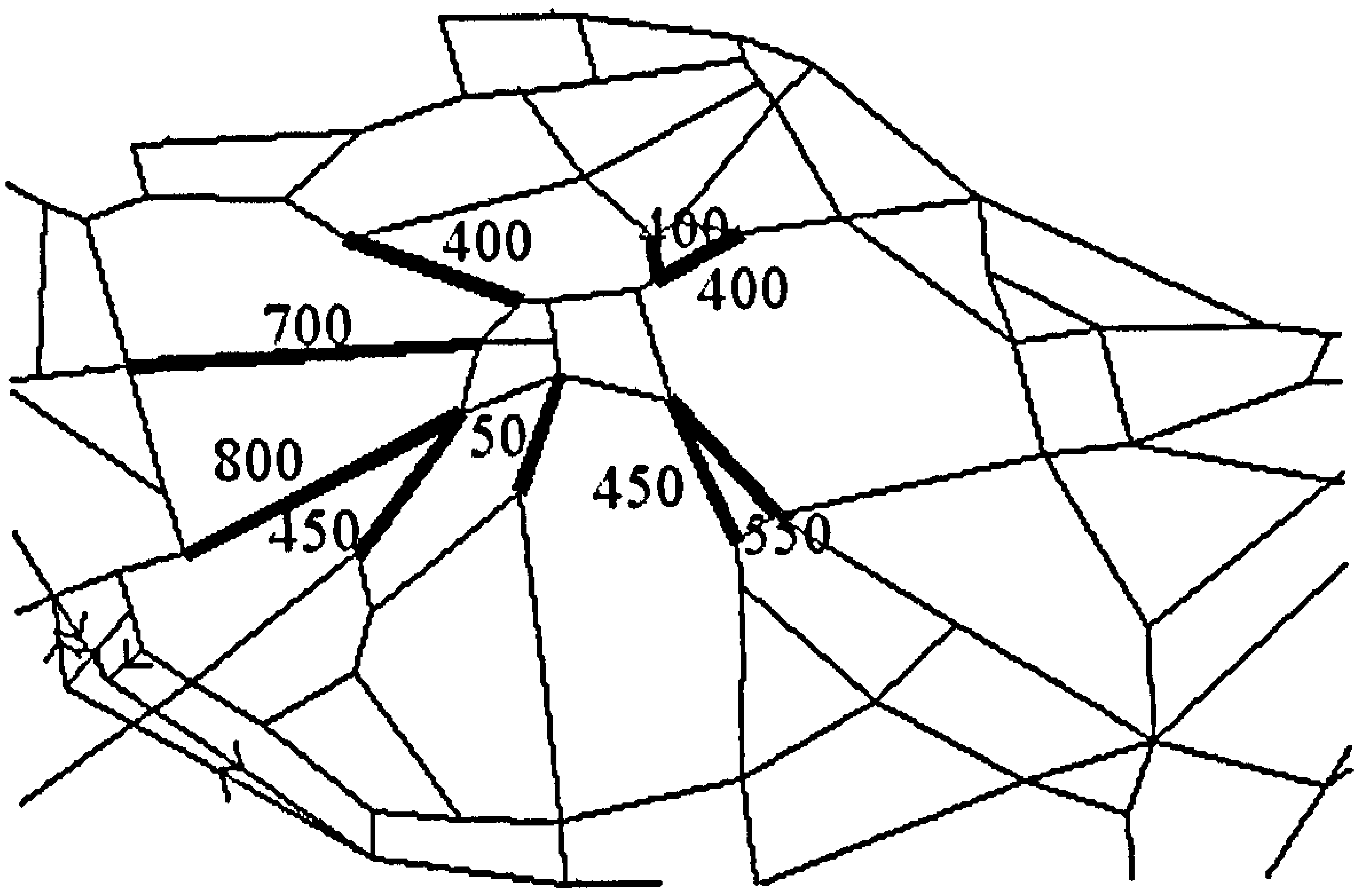


Figure 5-24 Optimal tolls for the inner1 cordon found by GA-CHARGE (highlighted links are tolled links)

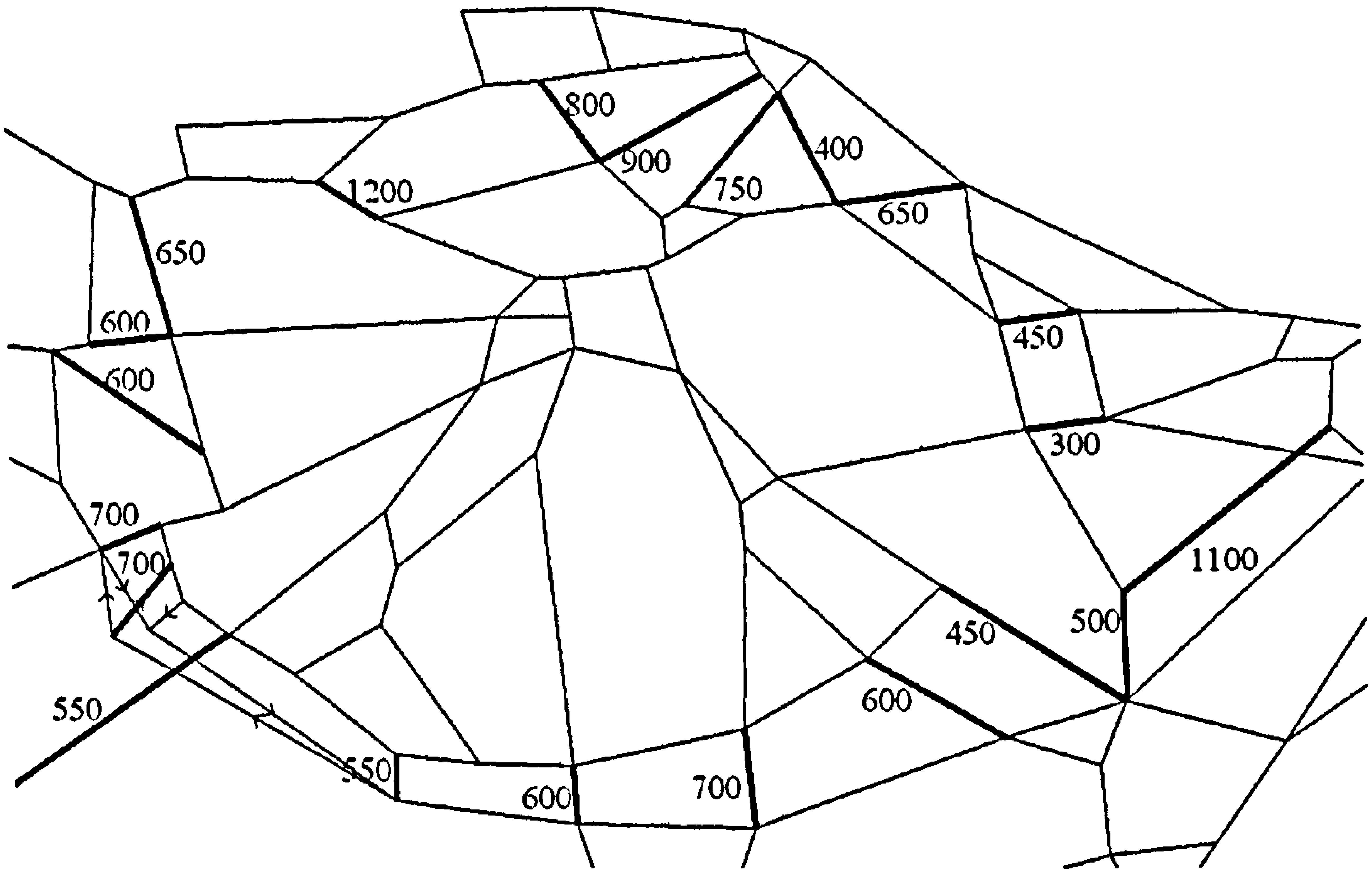


Figure 5-25 Optimal tolls for the outer cordon found by GA-CHARGE (highlighted links are tolled links)

Overall, GA-CHARGE is found to be a very effective algorithm for solving the COTP both with small and large networks. The only concern found during the experiment is the amount of time required by the GA process. For the tests of the Edinburgh network, each run of the equilibrium assignment and the calculation of the fitness of the chromosome requires approximately 3 CPU-seconds. Thus, GA-CHARGE requires at least 8 CPU-hr to achieve the solution presented earlier for the outer cordon problem of the Edinburgh network. These tests were conducted on the personal computer with the CPU of 1.9 GHz and 256 MB of RAM.

Obviously, some of the solutions tested during the GA process are repeated. The inclusion of the procedure to detect the duplicated tests can possibly decrease the run-time of the algorithm significantly. In addition, the rapid development of the performance of the computer can also make the application of GAB algorithm to a larger network less time-consuming in the near future.

5.7 SUMMARY

This chapter reviewed the optimisation algorithms for solving the optimal road pricing design problem. Two parallel reviews were presented from the transport and optimisation perspectives. From the transport perspective, eight different categories of methods were discussed including (i) iterative optimisation approach, (ii) linearization approach, (iii) sensitivity analysis, (iv) inverse optimisation, (v) minmax/maxmin based optimisation, (vi) smoothing approach and reduced UE condition, (vii) cutting plane algorithm, and (viii) meta-heuristic optimisation approach. Similarly, four different groups of the optimisation strategies for MPEC were presented including (i) disjunctive nonlinear programming based algorithm, (ii) penalty and interior point based method, (iii) implicit programming based method, and (iv) nonlinear optimisation based method. All of these methods exploit different characteristics of the MPEC and transform MPEC to an equivalent single level optimisation problem in different manners.

After an extensive review, three possible methods were proposed for solving COTP. The first method is the smoothing algorithm approach. The inherit complementarity condition of the MPEC problem induces a number of problems with the conventional nonlinear optimisation method including the non-differentiability of the CP condition and the violation of the constraint qualification. The strategy adopted in the smoothing algorithm is to replace the complementarity condition by its equivalent perturbed merit function. In particular, the perturbed Fischer-Burmeister (F-B) function is adopted. The special property of the perturbed merit function or in particular the perturbed F-B function is that (i) it is continuously differentiable everywhere when the perturbation parameter is greater than zero and (ii) it is equivalent to the CP condition when the perturbation parameter is reduced to zero. Thus, by replacing the CP condition with the perturbed F-B function, a sequential optimisation algorithm can be developed to tackle the MPEC as a conventional nonlinear optimisation problem. The smoothing algorithm can be applied to both the UE condition in path-based or multicommodity link flow forms. The smoothing algorithm was tested with a five-link network with two O-D pairs. The algorithm successfully solved the problems tested (five single tolled link and three two tolled link cases).

An alternative method based on the cutting plane algorithm was also proposed in the chapter. The improved cutting plane algorithm exploits the structure of the feasible region of the link flows (which is a polyhedron). A number of extreme points will be created iteratively during the optimisation process and added to the equilibrium constraint that is expressed in a VI form. The Master Problem is thus a relaxed MPEC problem with the VI condition restrained by a number of extreme points generated so far. The optimal solution from the Master Problem is then used to find the relevant cut (shortest path) by solving the Sub Problem. The optimisation iterates between these two problems until convergence. The additional improvement made in the algorithm proposed in this chapter is the representation of the feasible flow region in the Master Problem as a convex combination of a set of extreme flows in the same manner with the representation of the feasible flow for VI. The cut generated in the Sub Problem is also inserted into the convex combination of the feasible flow. With this reformulation, the number of variables can be decreased significantly. The algorithm was tested with the Pacman network and found to be very efficient.

The last algorithm proposed for solving the COTP is the Genetics Algorithm based method, GA-CHARGE. GA does not relies on derivative information of the optimisation problem, and hence can be applied to MPEC problem directly. GA-CHARGE was tested with the Pacman and Edinburgh networks. GA-CHARGE was proved to be effective in solving the optimal toll problem from the results with the Edinburgh network. However, the main problem with these methods is the computational time. This issue can be resolved with the introduction of 'memory' into the GA process, in which the duplicated solution will not require its fitness to be evaluated, hence reducing the number of UE assignments required.

Overall, the improved cutting plane algorithm was found to be the most efficient algorithm for solving the continuous optimal toll problem. However, the method has not been tested with a large scale problem (e.g. Edinburgh network) and this should be considered as an important future research task. The GA based method was found to be a very simple and effective algorithm for solving the optimal toll problem. Although, the GA based method required a substantial higher amount of time to solve the same problem compared to the improved CPA, additional improvement of the algorithm is envisaged to help reduce the time required. The merit function based algorithm was found to have a problem with the number of variables involved in the optimisation problem. This creates a problem when the algorithm is applied to a large scale problem. With the simplicity and flexibility of GA to handle different forms of problem (e.g. involving a more complex constraint or combinatorial problem), GA was considered as the most appropriate algorithm for the development of techniques for designing the practical road pricing scheme in the following chapters.

The methods reviewed and proposed in this chapter aim mainly to tackle the theoretical design of a road pricing scheme. None of the practical design criteria found earlier in Chapter 2 and Chapter 3 were included in the optimisation. In fact, the more theoretical algorithms like the smoothing based method or the improved cutting plane algorithm may not be able to handle some of the practical constraints. Thus, in the next chapters GA is chosen as the main optimisation mechanism. The next chapter deals with the inclusion of the topological constraint of the road pricing scheme design.

CHAPTER 6 OPTIMAL CHARGING CORDON DESIGN

What has now been proved was once an imagination.

6.1 INTRODUCTION

In Chapter 3, several practical design criteria of a road pricing scheme were identified. Some key characteristics of a practical road pricing scheme found are related to the topology of charging points. As mentioned earlier in Chapter 3, it is widely believed that a charging cordon is the most user-friendly charging system due to its simplicity and clarity to the road users. This is considered to be a key ingredient for promoting public acceptability of the charging scheme. A charging cordon is a set of tolled links surrounding a designated area so that all trips entering or passing through that area are tolled. Various current real world implementations of the scheme including the ERP scheme in Singapore, and Norwegian toll rings (in Oslo, Trondheim, and Bergen), and the congestion charging scheme in London.

The common practice in designing a charging cordon scheme documented in various real-world cases and desk studies (May, 1975; Holland and Watson, 1978; Transpotech, 1985) is to apply a judgmental approach or a *'trial and error process'* to seek appropriate toll locations and toll levels. The trial and error process normally starts by defining a set of possible charging cordons and their associated common toll levels. Then, each option is tested with some traffic modelling software. The benefit and cost of each scheme is then calculated from the modelling output and the best scheme from those considered is chosen.

The sub-optimality of this judgmental design approach is well addressed in May *et al* (2002) in which the benefits of cordons based on the judgmental and optimal designs are compared. They also suggested that the application of a theoretical approach could improve the benefit of the judgmental design significantly. One possibility is to directly optimise the location of charging cordons using optimisation theory. The motivation of this chapter is to develop an optimisation method to find optimal locations specifically for charging cordons.

Recently, there have been an increasing number of attempts to develop analytical approaches to tackling the optimal toll level and location problems(Mun *et al*, 2001; Hearn and Yildirim, 2002; Hyman and Mayhew, 2002; May *et al*, 2002; Verhoef, 2002; Ho *et al*, 2004; Lawphongpanich and Hearn, 2004; Shepherd and Sumalee, 2004; Zhang and Yang, 2004). A more extensive review on this topic was presented earlier in Chapter 4 and will not be reiterated here. Briefly, most of the literature concentrates on deriving the solution of the problem formulated in equation 4.1 in Chapter 4 which is of theoretical interest. Only some of the research considered the requirement for a charging cordon. Mun *et al* (2001), Hyman and Meyhew (2002), and recently Ho *et al* (2004) developed approaches to define the optimal location of a charging cordon with a non-network based modelling framework (using either a continuum traffic model or strategic model).

However, the omission of network representation in the model may cause a sub-optimal design of the cordon location. There are two particular reasons; firstly congestion is a local issue which happens mostly on links rather than in wide areas; secondly the analysis of charging scheme without a network representation may underestimate the impact of re-routing (Milne, 1997). Thus, it is considered more appropriate to tackle the cordon design problem under the framework of network modelling.

Zhang and Yang (2004) developed an approach based on *'genetic algorithms'* (GA) to find an optimal closed charging cordon with an optimal uniform toll under a detailed network representation. The method proposed later on in this chapter was independently developed based on a similar principle, with GA being utilised as the main optimisation mechanism.

The difference between the method in this chapter and the one in Zhang and Yang (2004) is largely in the chromosome design (which is the proxy for a charging cordon in GA). Their chromosome represents the status of nodes, identifying whether they are tolled or un-tolled. From a set of tolled and un-tolled nodes, the links starting from one of the un-tolled nodes and ending at one of the tolled nodes are defined as tolled links.

They also proposed a method to verify whether a set of tolled links forms a charging cordon. However, this chromosome design does not ensure that after the crossover or mutation operators are applied the new chromosome will represent a charging cordon. On the other hand, the method in this paper ensures that the new chromosomes conserve the closed charging cordon formation after the GA operators are applied.

The GA based method proposed in this chapter is termed GA-AS. The chapter is structured into five further sections. The next section presents an innovative approach to defining a set of charging cordons. The third section introduces the concept of GA and explains the method developed, named GA-AS. Section four then presents an extension of the method to the design of a multiple charging cordons scheme. Section five investigates the influence of different GA parameters by testing GA-AS with the Edinburgh network. The last section summarises the chapter.

6.2 OPTIMAL CHARGING CORDON DESIGN PROBLEM

The problem discussed in this chapter is to find the optimal location of tolled links forming a charging cordon or multiple charging cordons, with their optimal toll levels, designed to maximise or minimise a selected objective function (e.g. social welfare improvement, total travel time, accessibility, or local environmental improvement). The formulations of these objectives were discussed earlier in Chapter 4.

In order to predict the responses of travellers to the toll imposed, an assumption is made on how travellers in the traffic network behave. The well-known *'Wardrop's user equilibrium condition'* (Wardrop, 1952) as explained earlier in Chapter 4 is used to represent travellers' behaviour. The next section mathematically formulates the problem. The problem formulation turns out to be a special case of the mixed 1-0 optimal road pricing design problem described in Chapter 4.

6.2.1 Formulating the road pricing design problem

Recall the formulation of the mixed 1-0 road pricing design problem[13]:

---

[13] See notation in Chapter 4.

$$
\begin{aligned}
&\max_{(\tau,\varepsilon,\mathbf{v},\mathbf{d})} \psi_1(\tau,\varepsilon,\mathbf{v},\mathbf{d}) \\
&s.t. \\
&0 \le \tau_a \le \varepsilon_a \cdot \bar{\tau}_a \qquad \text{for } \forall a \in A \\
&(\mathbf{v},\mathbf{d}) \to sol\big(VI\big(F(\mathbf{v},\mathbf{d},\tau,\varepsilon),\Omega(\tau,\varepsilon)\big)\big)
\end{aligned}
\qquad \text{Equation 6-1}
$$

The objective of optimal road pricing design adopted in this chapter is to maximise the '*net total benefit*' which is the social welfare benefits minus the costs of the road pricing system (see Section 4.4 for the mathematical formulation). Note that the choice of the objective function used does not affect the ability of the method proposed in this chapter to solve the problem. This is because GA does not require any information about the gradient of the objective function used. Thus, GA can be applied to any problem as long as the objective function adopted can be evaluated.

An additional constraint added in this Chapter is that the set of tolled links selected must form a charging cordon (the exact definitions of a charging cordon and multiple charging cordon scheme are given later on). Let Θ be the set of all possible closed charging cordons. A feasible combination of tolled links for the optimisation problem must be one of the members of this set, $\varepsilon \in \Theta$:

$$
\begin{aligned}
&\max_{(\tau,\varepsilon,\mathbf{v},\mathbf{d})} \psi_1(\tau,\varepsilon,\mathbf{v},\mathbf{d}) \\
&s.t. \\
&0 \le \tau_a \le \varepsilon_a \cdot \bar{\tau}_a \qquad \text{for } \forall a \in A \\
&\varepsilon \in \Theta \\
&(\mathbf{v},\mathbf{d}) \to sol\big(VI\big(F(\mathbf{v},\mathbf{d},\tau,\varepsilon),\Omega(\tau,\varepsilon)\big)\big)
\end{aligned}
\qquad \text{Equation 6-2}
$$

This problem can be categorised as a Mathematical Program with Equilibrium Constraints (MPEC) (Luo *et al*, 1996) as described in Chapter 4. There are various problematic characteristics of MPEC, e.g. non-convex feasible regions, non-convex objective functions, and non-smooth objective functions. Solving the MPEC problem is a very complicated task let alone the combinatorial nature of the problem stated in Equation 6-2. In addition, it is infeasible to handle the constraint on the topology of tolled links (due to its complexity) with the traditional optimisation algorithms.

There does not exist any derivative-based optimisation algorithm that is capable of solving the problem stated in Equation 6-2. In this chapter, the idea to tackle this problem is to utilise the flexibility of GA in dealing with the constraint of a closed cordon and complementarity constraint representing the UE condition (see Section 4.2.2). GA is first used to produce a set of charging cordons. The users' responses, following users' equilibrium, to the cordon toll are then evaluated by running an appropriate traffic assignment model. The fitness of each closed cordon will be evaluated from the users' responses. GA will then iteratively evolve the set of charging cordon solutions until reaching the predefined stopping criteria (e.g. maximum number of generations within GA).

In this chapter, a discrete uniform toll is assumed to be a design specification. In finding an optimal uniform toll for each cordon, each charging cordon will be evaluated in terms of its net total benefit with the different predefined toll levels. The toll level producing the highest net total benefit will be accepted as the optimal uniform toll for that cordon. In the next section, the concept of the branch-tree framework developed to represent a charging cordon is described; this is the core of this method.

6.3 BRANCH-TREE FRAMEWORK

6.3.1 Definition of a charging cordon

Definition 6-1: *Charging cordon.* A charging cordon is a set of tolled links surrounding a designated area so that all private car users travelling to the destination inside the area or through the area will be charged. A definition of a closed cordon in the context of graph theory is that all paths from all zones outside the cordon passing through the nodes inside the cordon must be tolled at least once on a link related to those paths.

Zhang and Yang (2004) defines a charging cordon based on the concept of a cutset in graph theory. It is worthwhile to reiterate the concept proposed in their paper in order to distinguish the differences between the methods developed in this chapter and their method.

Definition 6-2: *Cutset.* For a directed graph $G = (N, A)$, a cutset of G is a subgraph consisting of a minimal collection of links whose removal reduces the rank of G by one.

Definition 6-3: *Component.* A component of a graph is a connected subgraph containing the maximal number of edges. If a graph is not strongly connected, it must contain a number of components. Otherwise, a graph has only a single component.

Assumption 6-1: In this chapter, only a complete directed graph is considered which means there is only one component of the network. In other words, there is at least a path connecting each pair of nodes in the network.

Definition 6-4: *Incident matrix.* An incident matrix (**B**) with the dimension of $N \times A$ is a representation of a graph (G) where the elements of the matrix ($b_{n,a}$) are defined as follows:

$$b_{n,a} = \begin{cases} 1 & \text{if node } n \text{ is the origin node of link } j \\ -1 & \text{if node } n \text{ is the destination node of link } j \\ 0 & \text{otherwise} \end{cases}$$

Lemma 6-1 [Theorem 2.1: Chen, 1997]: For a graph with N nodes and β components, the rank of the graph (γ) is defined as the number $\gamma = N - \beta =$ the number of linearly independent rows or columns of the incident matrix (**B**) of G.

Theorem 6-1 [Zhang and Yang, 2004]: Given a graph representing a road network and a cutset, tolled links forming a charging cordon are those whose original node is in the component outside the cutset and destination node is in the component inside the cutset.

Proof. From Assumption 6-1, the graph considered is complete, so it only has one component. A set of links forming a cutset reduces the rank of the matrix by one and only one (due to Definition 6-2) which means it will increase the number of components of the graph by one and only one. This is true because the number of nodes in the graph before and after applying the cutset is constant and from Lemma 6-1 $\gamma = N - \beta$. Thus, the links forming a cutset will split the network (graph) into two components. By putting the tolls on all links heading toward the component inside the cutset all paths entering this area are charged which satisfies the definition of a closed charging cordon in Definition 6-1. □

Remarks: A cutset may not form a desired (practical) charging cordon, i.e. a cutset may not separate the desired charging area from the rest of the network (see the requirement in Definition 6-1). An additional condition must be made regarding the set of nodes that must be inside one of the components defined by a cutset (indicating a desired charging area), referred to as charged nodes. From Figure 6-1, suppose that the desired charging area is the black node in the centre of the figure and there is no constraint on the set of charged nodes. From this condition, cutset 1 satisfies the standard condition for a cutset but it does not form a desired charging cordon surrounding the desired charging area.

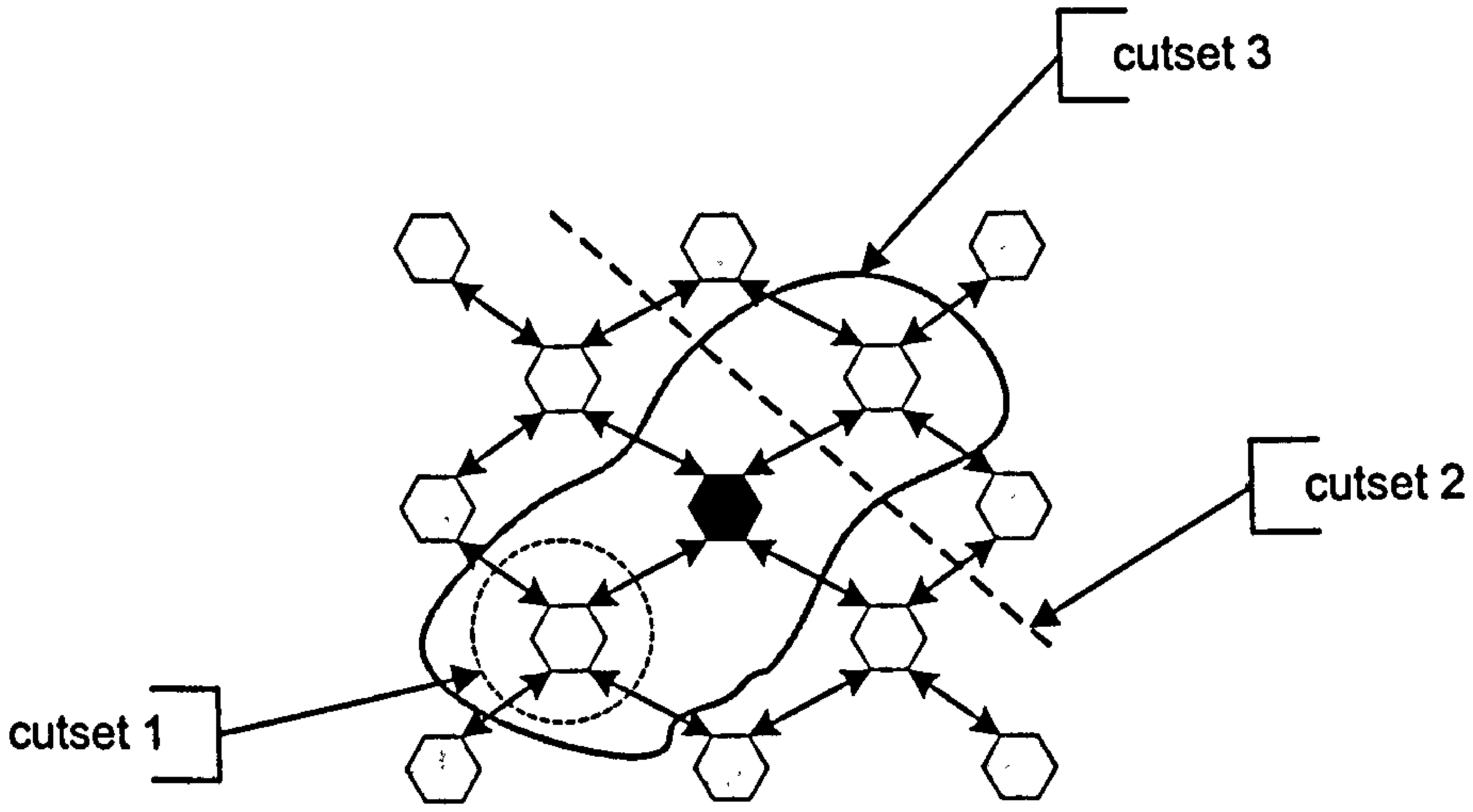


Figure 6-1 Example of drawback of cutset definition

However, although the condition on the charged nodes is imposed (i.e. the central node is defined as the charged node), cutset 1 and cutset 2 shown in Figure 6-1 still satisfy the condition for a cutset and charged node. The remedy for this problem is to include an additional constraint on the other set of nodes (referred to as boundary nodes). Boundary nodes and charged nodes must be in different components of the graph. From Figure 6-1, if an additional constraint on boundary nodes is imposed, neither cutset 1 nor cutset 2 satisfies this condition since some of the boundary nodes and charged nodes are in the same components of the graph. Cutset 3, on the contrary, satisfies this condition and this cutset forms a charging cordon according to Definition 6-1.

The method proposed in Zhang and Yang (2004), thus, could experience this drawback. In addition, the GA chromosome design in their method may not form a new charging cordon after the GA operators are applied to the chromosomes. The branch-tree approach proposed next overcomes these two drawbacks. The branch-tree concept does naturally conform to the requirement in Definition 6-1. In the next section, the concept of a branch-tree approach for defining a charging cordon is explained.

6.3.2 Nomenclature of the branch-tree concept

Let *G = (N, A)* be a directed graph representing an urban traffic network where *N* and *A* are a set of nodes and links respectively in the graph. A link is defined by two nodes, *i* and *j* where $i \neq j; i, j \in N$.

Definition 6-5: *Set of all preceding nodes.* If the direction of a link is from *i* to *j*, *i* is termed *'the preceding node'* of *j*. The number of preceding nodes for each node *j* can be more than one. The set $\Xi_j = \{i | i \text{ is the preceding node of } j\}$ is defined to be the set of *'all-preceding nodes'* of node *j* where $|\Xi_j|$ is the size of the set (the total number of preceding nodes of node *j*).

Definition 6-6: *Node degree.* The degree of a node is the number of children or preceding nodes of that node in the branch-tree.

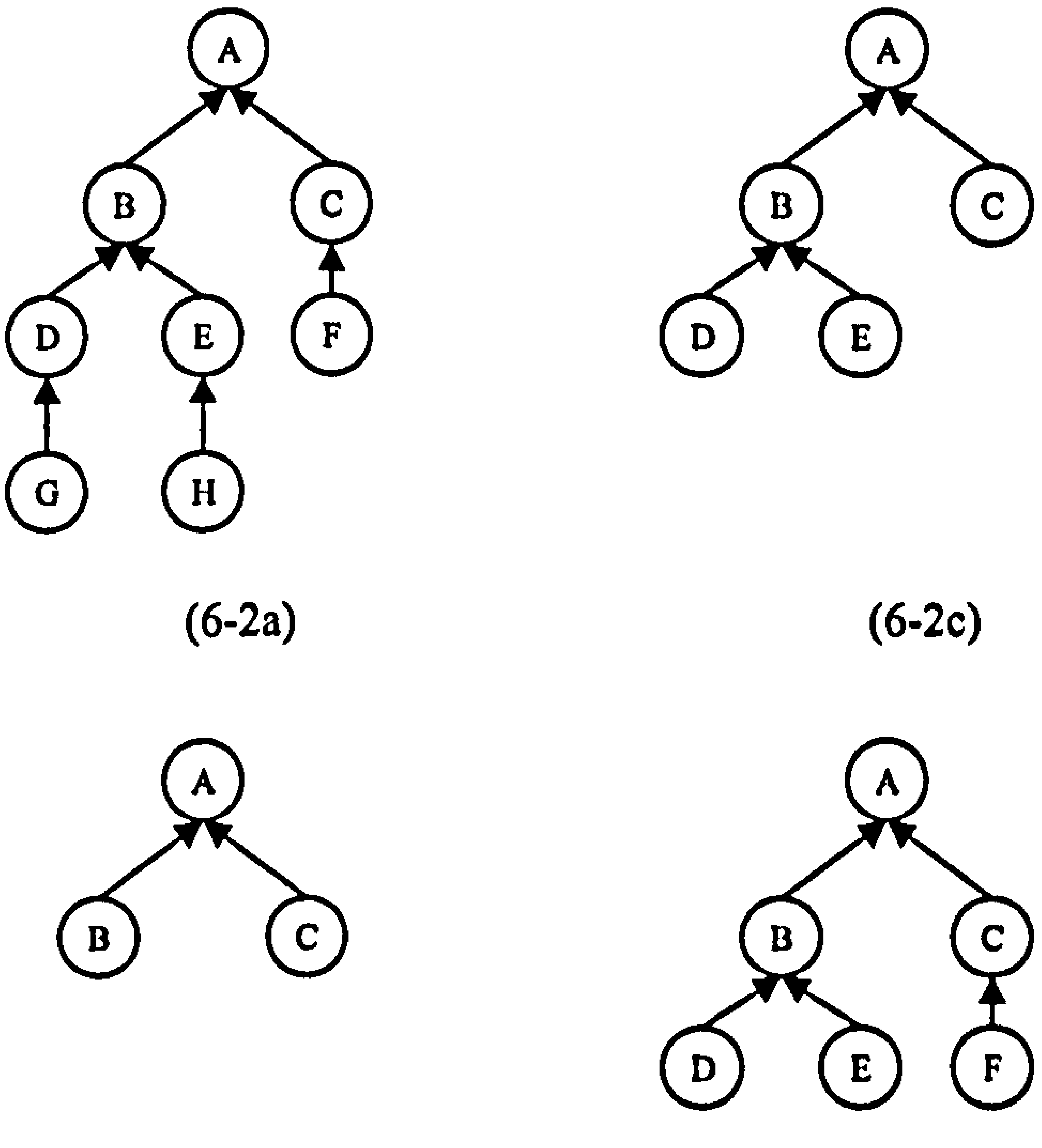
Definition 6-7: *Leaf node.* A node with degree of zero is called a *'leaf node'*.

Definition 6-8: *Branch-tree* and *Branching process.* $\beta_r = \{(n, d)\}$ is a branch-tree, which is defined as a set of the pairs of $n, n \in N$ and *d* (degree of node *n*). *r* is the root node of this branch-tree $(r \in N)$. $\beta_r = \{(n, d)\}$ has to be created from the original graph *G*. Given *G*, a root node (*r*) is defined and then only the preceding nodes of *r* can be included into the branch-tree, i.e. *n* can be included into the branch-tree if and only if $n \in \Xi_r$ (See Definition 6-5). When node *n* is added into the branch-tree, the degree of node *n* is initially set as zero (since node *n* has no children nodes in the branch-tree at this stage). Node *n* can then be expanded by including its preceding nodes into the branch-tree. The set of preceding nodes of node *n* added into the branch-tree are referred as children nodes of node *n*. After adding the preceding nodes of node *n* into the branch-tree, the degree of node *n* will be changed from zero to the number of children nodes of node *n*. This is the process to expand the *'depth'* of the branch-tree which is referred to as the *'branching process'*. The branching process can be applied iteratively to other leaf nodes added into the branch-tree. Once the process terminates, $\beta_r = \{(n,d)\}$ is produced. Example 6-1 below illustrate the branch-tree structure and branching process.

Example 6-1 Illustration of the branching process

Figure 6-2a shows an example of a full traffic network (G). Assume that node A is defined as root node $\beta_A = \{(A,0)\}$. According to Definition 6-5 given earlier, nodes B and C are the preceding nodes of node A and they can be added into the branch-tree β_A. Figure 6-2b shows the branch-tree after applying the branching-process to node A (adding nodes B and C into the original branch-tree shown in Figure 6-2b). At this stage $\beta_A = \{(A,2),(B,0),(C,0)\}$. Notice that the degree of node A is changed from zero to two (see Definition 6-8 above). Nodes B and C are now the leaf nodes of the branch-tree β_A.

Assume that only node B will be expanded. Nodes D and E which are the children nodes of node B are added into the branch-tree during the branching process creating the new branch-tree as shown in Figure 6-2c. At this stage, $\beta_A = \{(A,2),(B,2),(C,0),(D,0),(E,0)\}$. Again, notice the change of the degree of node B from zero to two. If the branching process is terminated at this stage, the final branch-tree β_A is the one shown in Figure 6-2c. Applying the branching process to expand the depth of the branch-tree with different set of leaf nodes can create different branch-trees from the same graph (see Figure 6-2d for different branch-tree from the graph in Figure 6-2a where the branch-tree is also expanded at node C).



(6-2a) (6-2c)

(6-2b) (6-2d)

Figure 6-2 Example of branch-tree and branching process

Definition 6-9: *Sub-branch*. Recall that $\beta_r = \{(n,d)\}$ is a branch-tree. Given *n* which is one of the nodes in $\beta_r = \{(n,d)\}$, $\bar{\beta}_n \subset \beta_r$ is a sub-branch of β_r. Inside a branch-tree, a number of sub-branches can be defined. Assume that node *n* from a full branch-tree is selected, the sub-branch created from this node is the whole part of the branch-tree rooted from node *n*. Figure 6-3 shows an example of a sub-branch rooted from node B of the full branch-tree shown in Figure 6-2c. In this example node B from the branch-tree (β_A) shown in Figure 6-2c is selected as the root node of the sub-branch. $\bar{\beta}_B$, which is a sub-branch of the branch-tree, is thus the whole part of the branch-tree rooted from node B which are node D and E. Therefore, $\bar{\beta}_B = \{(B,2),(D,0),(E,0)\}$ and thus $\bar{\beta}_B \subset \beta_A$.

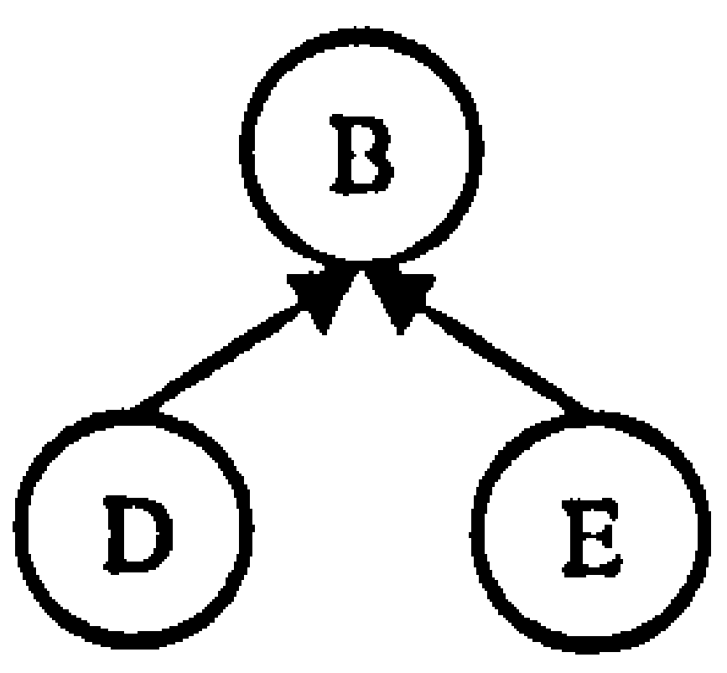


Figure 6-3 Example of sub-branch

Definition 6-10: *Branch.* Given a branch-tree $\beta_r = \{(n,d)\}$ and its associated original graph *G*, a branch is defined as a link $a \in A$ defined by a leaf node of $\beta_r = \{(n,d)\}$ as its origin node and a proceeding node of that branch node in $\beta_r = \{(n,d)\}$ as its destination node. For example, from the branch-tree shown in Figure 6-2d, the branches of this branch-tree are links D-B, E-B, and F-C.

6.3.3 Relationship between branch-tree and closed charging cordon

After introducing the necessary notation for the branch-tree concept, this section explains the relationship between the branch-tree concept and the closed cordon formulation.

Proposition 6-1: For a given branch-tree ($\beta_r = \{(n,d)\}$), the tolled links are defined by a set of branches in the branch-tree. These tolled links will form a closed charging cordon around the root note *r if and only if* all nodes in $\beta_r = \{(n,d)\}$ have either $d = 0$ or $d = |\Xi_n|$. This implies that in expanding a leaf node (using the branching process) all preceding nodes of that leaf node must be included into the branch-tree. This is exactly the process adopted in the example in Figure 6-2.

Proof. This proposition can be easily verified by considering the definition of a closed charging cordon mentioned earlier. By including all preceding nodes of a leaf node into the branch-tree, all links entering that leaf node will be defined as tolled links and hence all paths entering that (previously) leaf node are tolled. From a root node (*r*), by including all preceding nodes of the root node into the branch-tree all paths entering that root node are tolled. In expanding the leaf node *j* (which are all preceding nodes of node *r*), from the proposition all preceding nodes of node *j* ($\forall n | n \in \Xi_j$) must be included into the branch-tree. Note that now all paths entering node *j* must pass through one of the preceding nodes of node *j*. By tolling the links defined by node *j* and its preceding nodes all paths entering node *j* are tolled. Previously, all paths entering node *r* via node *j* are tolled by the tolled link (*r*, *j*). After expanding node *j*, the new set of tolled links ensuring that all paths entering node *j* are tolled and hence all paths entering node *r* via node *j* are still tolled. This proof can be expanded to the general case between node *j* and *j+1* where node *j+1* is one of the children nodes of node *j*, hence it proves that the condition of a branch-tree proposed is sufficient for defining a charging cordon according to Definition 6-1. □

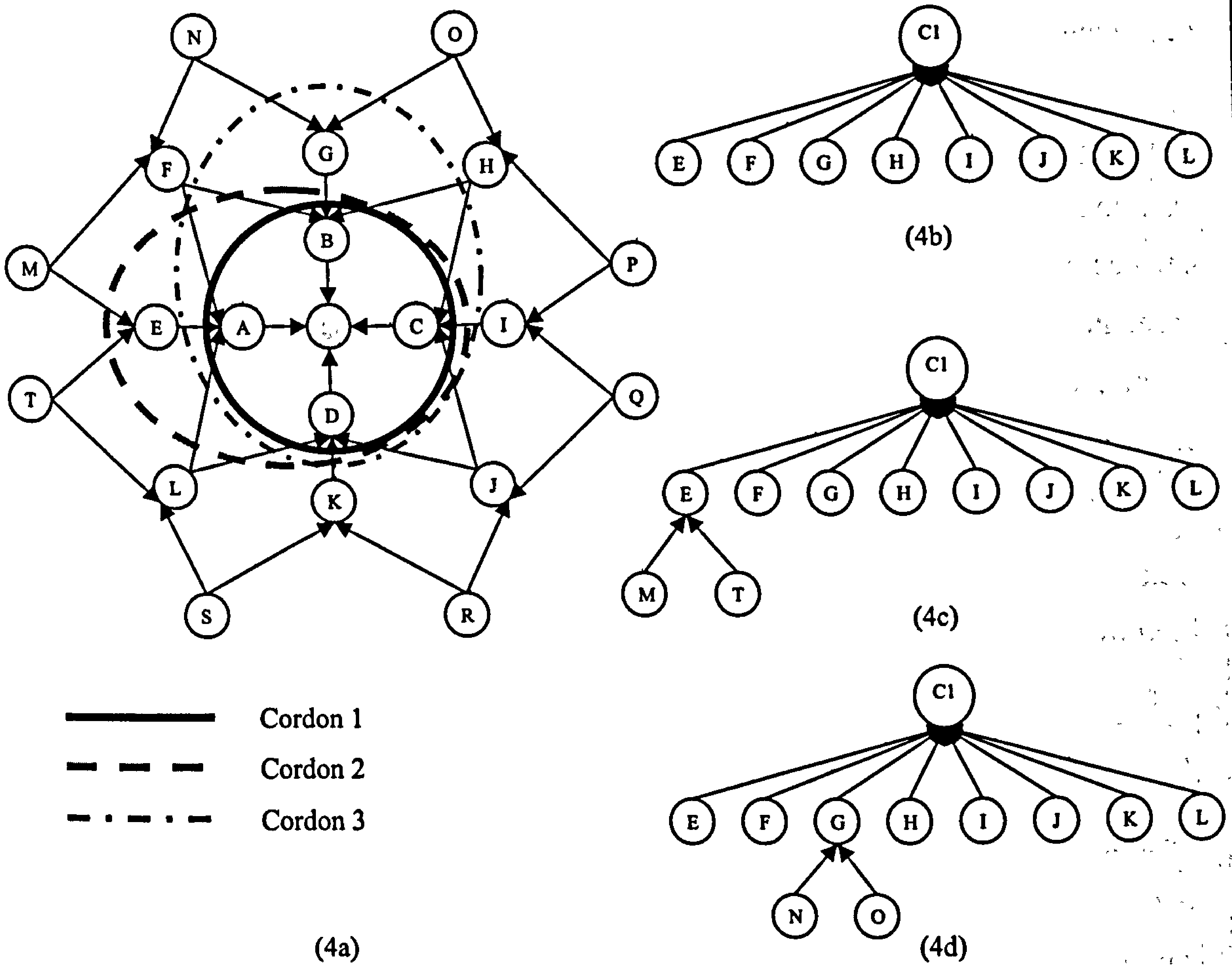
Before generating different charging cordons, a set of predefined links forming an initial charging cordon in the network must be defined. The starting node of each link will become a root node for the branch-tree. For instance, assuming that in a network the predefined cordon comprises five links, five branch-trees will be generated to define the set of charging cordons. These five branch-trees generated from the predefined cordon can be combined into a single global branch-tree with a given virtual root node reducing five branch-trees to a single branch-tree which represents a charging cordon.

Example 6-2 Relationship between branch-trees and charging cordons

Figure 6-4a shows the hypothetical network used for exemplification. The grey node is assumed to be the city centre which is the preferred tolled area. In a network, as mentioned earlier a set of links forming an initial closed cordon around the tolled area must be predefined. Assume that Cordon 1 in Figure 6-4a is defined to be the initial cordon. From this initial cordon, a virtual root node (name "C1") is defined for the branch-tree, and the first level nodes in the branch-tree are the preceding nodes of the links forming the initial cordon (Cordon 1). Figure 6-4b shows the branch-tree C1 representing the initial cordon.

Assume that node E is to be expanded to create a new cordon. The original branch-tree in Figure 6-4b is then expanded at node E, creating the new branch-tree as shown in Figure 6-4c. This new branch-tree forms Cordon 2 shown in Figure 6-4a. In Figure 6-4d, the original branch-tree is instead expanded at node G to produce Cordon 3. All three cordons in Figure 6-4a are closed cordons. The tolled links in each cordon are defined by the branches in the branch-trees. Note that nodes E-L, which are predefined by the user, are referred to as *'target nodes'* because they are the set of nodes where all paths passing through these nodes must be tolled. The branching process will be applied to each target node in turn to define the shape of the cordon, as illustrated in Figure 6-4.

This notation will be used in the algorithm for generating a closed charging cordon in the next section.



(4a)

(4b)

(4c)

(4d)

Figure 6-4 Relationship between branch-trees and charging cordons

Next, several issues related to the structure of a branch-tree and conventional rules for branching process will be discussed.

6.3.4 Dummy nodes in a branch-tree

For a tree, as defined in graph theory, there must be a *'unique path'* between the root node and each node in the tree (see Chen, 1997 for more detail). This original concept of a tree is not valid for a branch-tree built from a traffic network.

Definition 6-11: *Dummy node.* For a given branch-tree, a node with more than one acceding nodes in the branch-tree is referred to as a dummy node.

Example 6-3 Dummy nodes example

Figure 6-5 illustrates this problem where node H has two acceding nodes, i.e. nodes E and F, and hence there are two possible paths to traverse between nodes A and H in the branch-tree. Let $\bar{\beta}_B$ and $\bar{\beta}_C$ be two sub-branches of branch β_A in Figure 6-5. If $\bar{\beta}_B \cap \bar{\beta}_C \neq \varnothing$, then $\bar{\beta}_B$ and $\bar{\beta}_C$ are referred to as non-separable sub-branches. This means sub-branches $\bar{\beta}_B$ and $\bar{\beta}_C$ must have some overlapping parts (nodes) (see Figure 6-5). There must be at least one common node (θ) appearing in both sub-branches, $(\theta,d) \in \bar{\beta}_B \cap \bar{\beta}_C$. These nodes are named *'dummy nodes'* such as node H in Figure 6-5.

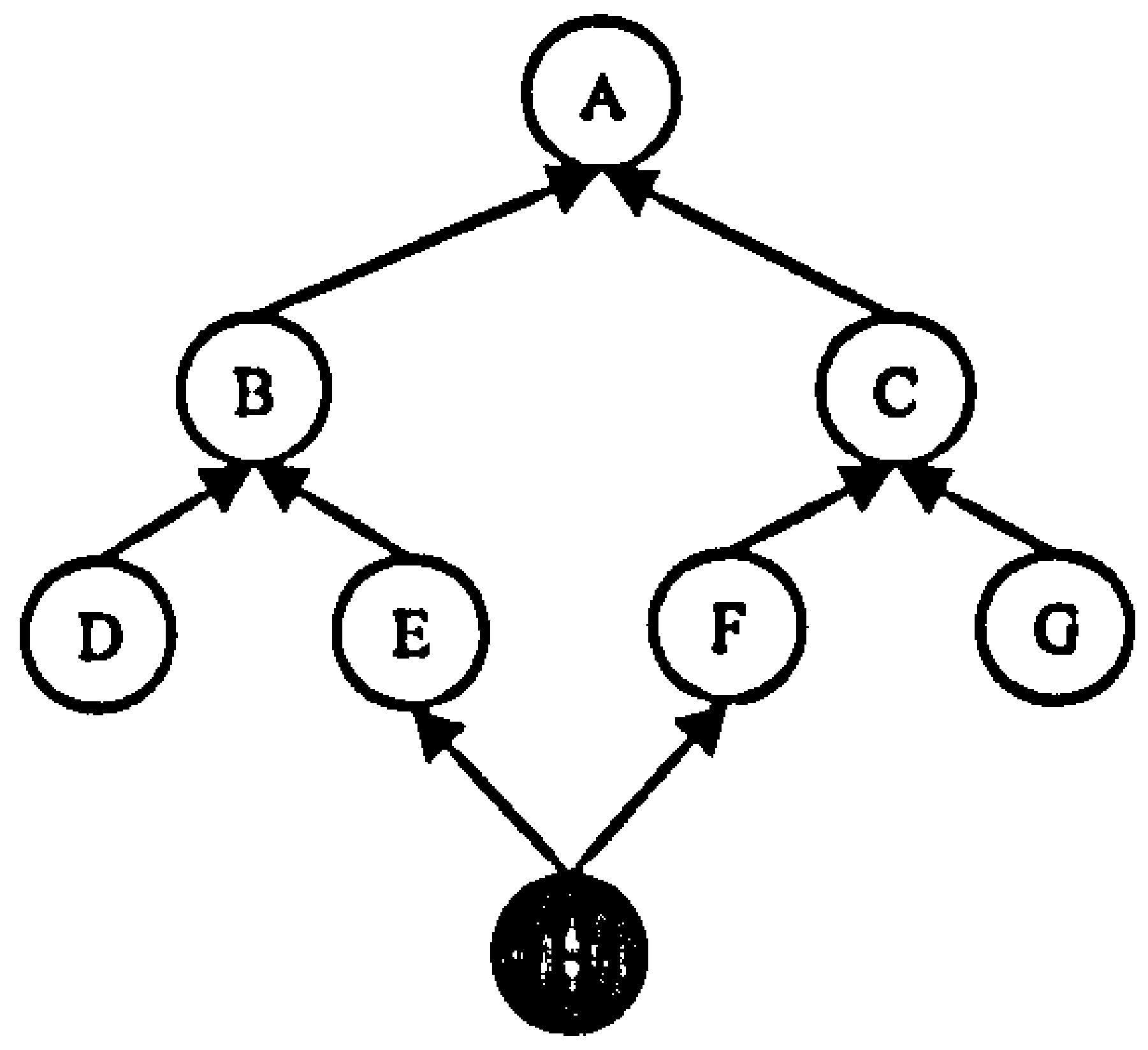


Figure 6-5 Dummy node in branch-tree

Dummy nodes will be expanded in only one sub-branch in order to avoid the inconsistency in the structure of the cordon. In order to do this, we establish rules for expanding the branch as follows:

Algorithm 6-1:

Step 1: Check if there is θ (dummy node) in the branch-tree, if true go to step (2), otherwise continue branching process as described in section 6.3.2.

Step 2: For each θ, check if θ in all sub-branches are leaf nodes; if so go to step (3). Otherwise go to step (4)

Step 3: Label θ as "Dθ" and put θ on the branch with zero degree as normal. The reason for re-labelling dummy nodes is to hint to the crossover and mutation process in GA that changing this node involves other dummy nodes in the branch. Finish the process.

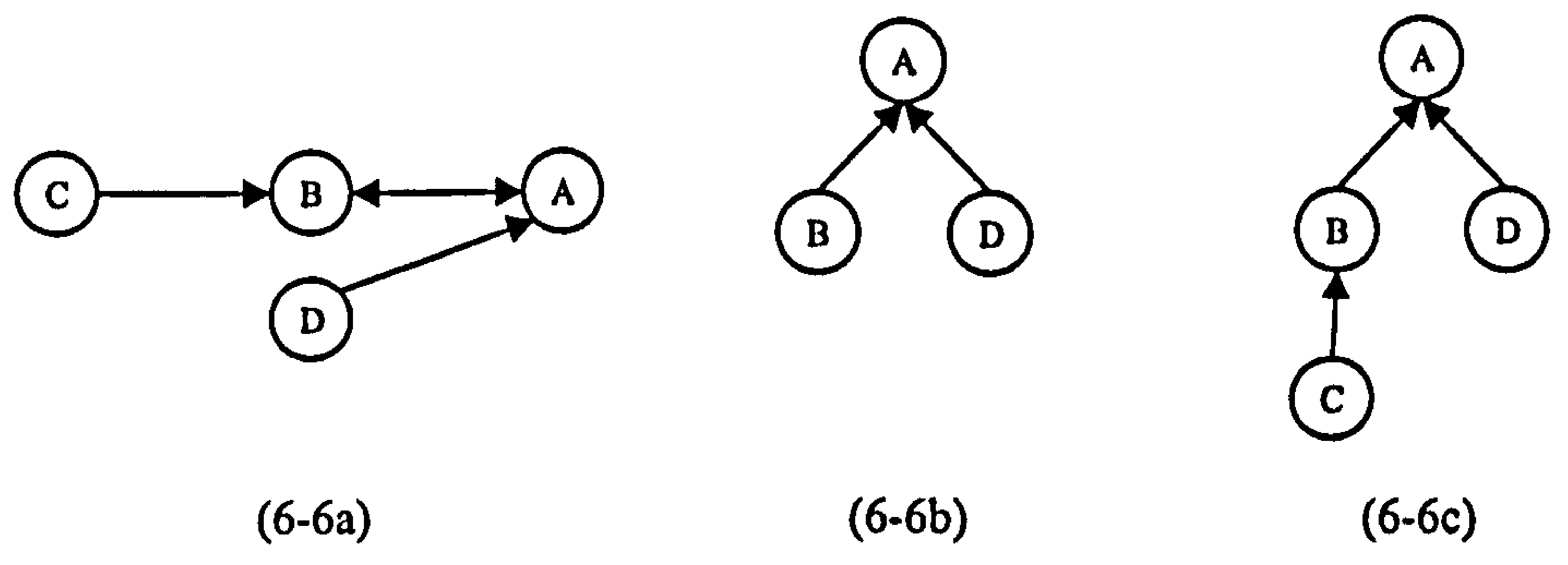
Step 4: Label node θ as "Dθ". Then put "D" as the node degree for the node in the sub-branch that will not be expanded. For the node in the sub-branch that will be expanded, operate the branching process as explained in Section 6.3.2.

6.3.5 Two-way link issue

Links and nodes in a graph, $G = (N, A)$, represent roads and junctions in a road network. For a two-way link defined by node *i* and *j*, node *i* is a preceding node of node *j* and vice versa. As mentioned, in the branching process a leaf node, *n*, can be expanded by including all nodes in Ξ_n into the branch-tree resulting in the change of the node degree from zero to $|\Xi_n|$. Assume that a leaf node *n* is expanded from node *j*. In fact, node *j* is one of the preceding nodes of node *n* due to its two-way link property. If node *n* is to be expanded, node *j* should not be included into the branch-tree again. If node *j* is included into the branch-tree, the toll will be imposed on the traffic coming from inside the cordon.

Example 6-4 Two-way links in general traffic network

Figure 6-6 is used to illustrate the issue of two-way links. Figure 6-6a is the road network used in this example. Note that the link between nodes A and B is a two-way link. The original branch-tree is shown in Figure 6-6b where the tolled links are links (B, A) and (D, A). Node A is used to represent the area inside the charging cordon defined by the branch-tree. Assume that the branch-tree is to be expanded at node B that has both nodes A and C as its preceding nodes. In this case, only node C will be added into the branch-tree, which represents link (C, B) (see Figure 6-6c). Node A will not be added into the branch-tree again. If node A were added into the branch-tree, link (A, B) would be defined as a tolled link. The toll on link (A, B) would then be imposed on the traffic coming from node A which is the traffic from inside the cordon. In the general process of building a branch-tree or in the branching process, there is a process to detect this two-way link issue to avoid tolling traffic coming out from inside the charging cordon.



(6-6a) (6-6b) (6-6c)

Figure 6-6 Two-way link issue for branch-tree formulation

6.4 APPLYING GA TO SOLVE THE OPTIMAL CLOSED CORDON

The optimisation algorithm developed for tackling the optimal cordon design problem is based on the idea of GA. The algorithm developed is termed GA-AS. Figure 6-7 depicts the overall process of GA-AS.

To solve the optimal closed cordon problem as stated in Equation 6-2, in the *'Initialisation'* stage GA-AS produces an initial population (the size of the population is defined by *'population numbers'*) representing a set of charging cordons encapsulated in the form of chromosomes (using the branch-tree formulation). Then, in the *'Evaluation'* process for each cordon an optimal cordon toll is found by testing it with different pre-specified tolls. Each toll level is implemented in the network, and then a traffic modelling software package (in this case, SATURN) is used to predict the responses of road users to the toll. The net total benefit from each toll level is calculated from the modelling outputs. The toll level producing the highest net total benefit is then selected as the optimal uniform toll for that cordon.

SATURN is a steady-state equilibrium assignment model that predicts route choice and traffic flows on a road network, based on the generalised costs of travel and takes account of delays due to capacity constraints (Van Vliet, 1982). It includes an assignment sub-model, which estimates driver route choices using Wardrop user equilibrium assumptions (Wardrop, 1952). In its conventional form, the model assumes fixed road travel demand. However, the capability exists to introduce variable demand through the SATEASY elastic assignment algorithm. This allows the representation of changes in demand that occur as a direct result of changes in the costs experienced on the road network (e.g. toll). It should be noted that any other traffic modelling software or even different equilibrium paradigms (e.g. stochastic equilibrium) could be used with the method described in this paper so long as it is able to produce the outputs required for the calculation of the net total benefit.

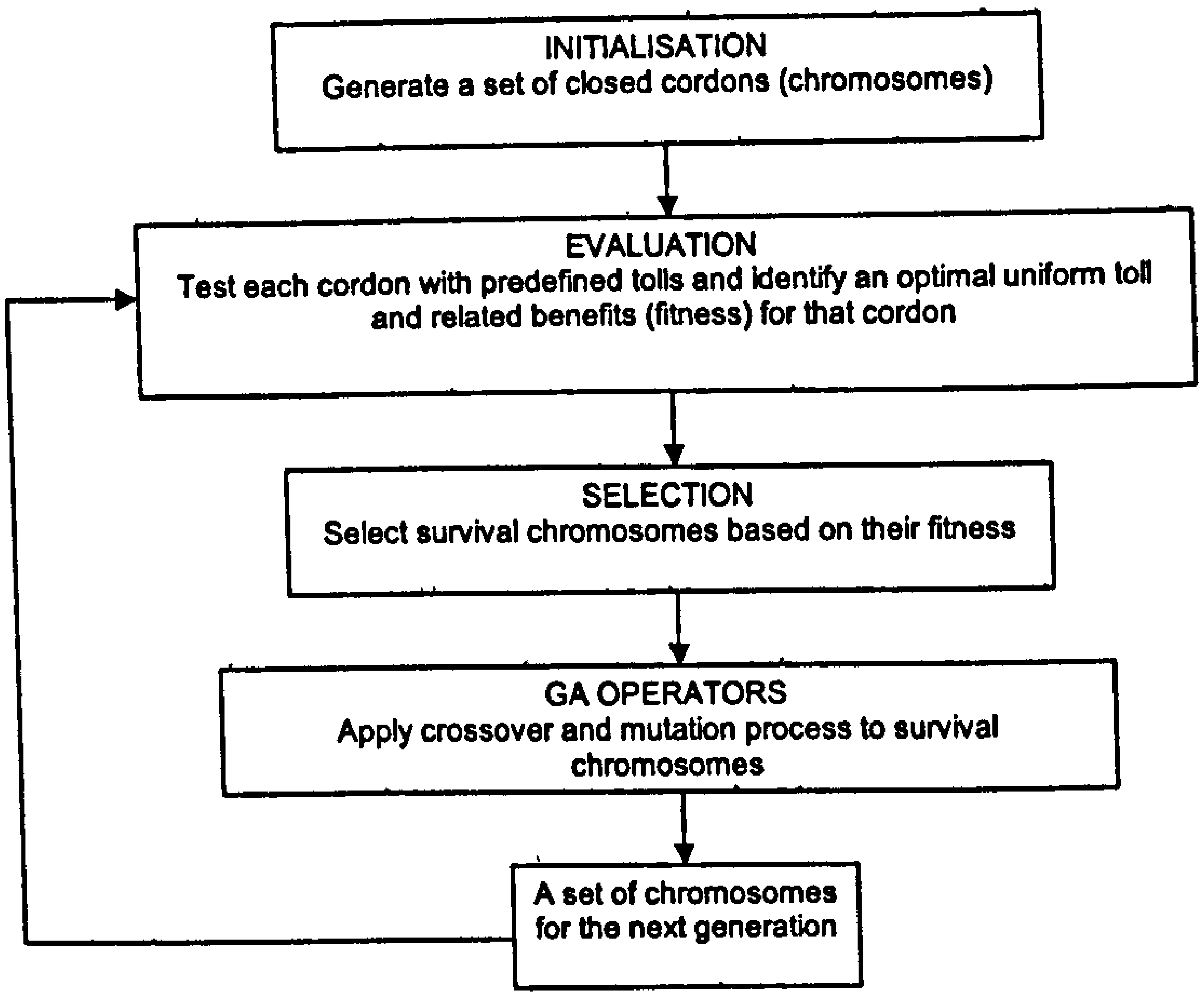


Figure 6-7 Overall process of GA-AS

After the optimal uniform toll for each cordon is found, the net total benefit (see the objective function in formula Equation 4-13 in Chapter 4) associated with the optimal toll is assigned as the fitness of that charging cordon (chromosome). Then, GA-AS, in the *'Selection'* process, selects the survival cordons (chromosomes) based on their fitness. The selected set of survival chromosomes are then crossed over and mutated sequentially (*'GA operators'*) to produce the set of chromosomes for the next generation. The process is terminated by the predefined number of generations (user input).

The chromosome design in GA-AS for closed cordons associated with the branch-tree concept is discussed next. The algorithm (based on recursive programming) to generate the initial set of charging cordons is then described. Next, the selection process, which is based on the method of roulette wheel linked with the linear ranking method proposed by Whitley(1989), is explained. Then, the two important genetic operators used in this method, i.e. crossover and mutation, are discussed.

6.4.1 Chromosome design

It is crucial to design a chromosome structure that is compatible with the structure of the branch-tree explained in the previous section. More importantly, the chromosome structure should be able to maintain the feasibility of the solution (in this case, the closed cordon format) even after applying the crossover and mutation process. As mentioned in the previous section, the members of a branch-tree have two key characteristics, i.e. their node numbers and degrees. Therefore, two strings, each a series of numbers or alphabets, will be used to represent a chromosome. The first string, called the node string, contains the node numbers of the branch-tree. The second string, called the degree string, contains the degree of each node in the corresponding column in the node string. Figure 6-8 shows an example of a chromosome.

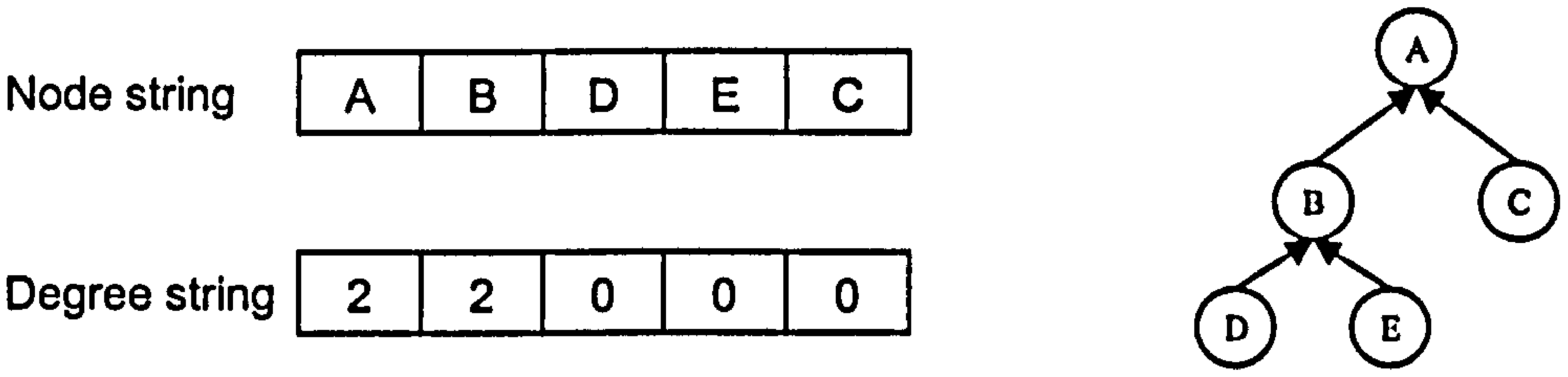


Figure 6-8 Chromosome structure for GA-AS

The node string tells us that this branch-tree comprises nodes A, B, D, E, and C. The degree string tells us that nodes A, B, D, E, and C have the degrees of 2, 2, 0, 0, and 0 in that order. This chromosome represents the branch-tree shown in Figure 6-2c. However, from the node and degree strings, the only information provided is which nodes are in this branch-tree and the degrees of each node. They do not provide any information about the connection of the nodes in the branch-tree. For example, how may one know that node B is the preceding node of node A without looking at Figure 6-2c? The answer is in the order of the nodes in the node string. The algorithm for encoding a chromosome from a branch-tree explains the meaning of the node sequence in the node string. The algorithm used to encode the branch-tree to node and degree strings is as follows:

Algorithm 6-2[14]:

Step 1: Set k to be the first column in node and degree string. Set j equals to 1. Put the root node and its degree into column k.

Step 2: Set P_j = a set of all preceding nodes of the node in k. Set $\overline{P}_j = \varnothing$.

Step 3: Check if $\overline{P}_j = P_j$. If so, set $j = j-1$ and then go to step 7. Otherwise go to Step 4.

Step 4: Set $k = k+1$.

Step 5: Set n_j to be a node in P_j but not in $\overline{P}_j$. Put n_j into the k column of node string

Step 6: Put the degree of n_j into the k column of degree string. Add n_j to $\overline{P}_j$. If degree of n_j is equal to zero, go to Step 3. Otherwise set $j = j+1$ and go to Step 2.

Step 7: Check if $j = 0$. If so, stop. Otherwise, go to Step 3.

Example 6-5 Demonstration of encoding a branch-tree into a chromosome format for GA-AS

From Figure 6-8, first node A is put into the first column of the node string and its degree (2) is put into the first column of the degree string. Then set $P_1 = \{B, C\}$ which is the set of all preceding nodes of node A. In Step 5, node B is picked and placed in the second column of the node string. In Step 6, the degree of node B, which is two, is placed into the second column of the degree string and node B is added into $\overline{P}_1$. Since the degree of node B is not equal to zero, j is set equal to 2 and we then go to Step 2. In Step 2, we set $P_2 = \{D, E\}$ and in Step 3 node D is picked and put into the third column of the node string. Also, the degree of node D, which is 0, is put into the third column of the degree column. Node D is also added into the set $\overline{P}_2$. Since the degree of node D is zero, the process moves straight to Step 3.

In step 3, since $\{D, E\} \neq \{D\}$, node E is picked and the same process is repeated as node D (the degree of node E is also null). The process returns to Step 3 again, and this time $P_2 = \overline{P}_2 = \{D, E\}$. Thus, j is set to 1, and the process moves to step 7 and then step 3 sequentially (since $j \neq 0$). Since $\{B, C\} \neq \{B\}$, node C is picked and put in the fifth

---

[14] The programming approach adopted to code this algorithm is the "recursive" program.

column. Again, its degree, which is zero, is put in the fifth column of the degree string. Now, $\bar{P}_1 = \{B,C\} = P_1$. Therefore, in Step 3, *j* is set to zero and the process is terminated.

6.4.2 Initialisation

In this stage, a number of charging cordons are randomly generated. The number of cordons generated in the first generation of the GA is controlled by the pre-defined population numbers. The generation of a closed cordon is designed as a random process. The variable '*Prop*' is defined as the probability of a node to be expanded (recall the definition of branching process explained earlier). The user must also define the initial closed cordon by defining a set of tolled links. The preceding nodes of these links will become a set of target nodes (see Figure 6-4). The initialisation process is as follows:

Algorithm 6-3:

```
For each target node defined by the user
    Step 1: Set $k$ to be the first digit in the node and degree string. Set $j$ equal to 1. Put the root node and its degree into column $k$.
    Step 2: Set $P_j = \Xi_{n_k}$ (a set of all preceding nodes of the node in $k$). Set $\bar{P}_j = \varnothing$.
    Step 3: Check if $\bar{P}_j = P_j$. If so, set $j = j - 1$ and then go to step 9. Otherwise go to Step 4.
    Step 4: Set $k = k+1$.
    Step 5: Set $n_j$ to be a node in $P_j$ but not in $\bar{P}_j$. Put $n_j$ into the $k$ column of the node string
    Step 6: Use algorithm 6-1 to check dummy nodes and check the issue of two-way links
    Step 7: Add $n_j$ to $\bar{P}_j$. If $n_j$ is a dummy node or representing the link in the opposite direction (two-way link issue), then go to Step 3. Note that if $n_j$ is ignored due to the two-way link issue, remove that node from $|P_j|$. Otherwise, go to step 8.
    Step 8: Generate a random number (rand). If rand > Prop, then put $|P_j|$ as the degree of $n_j$ into the $k$ column of the degree string; and then set $j = j+1$ and go
```

to Step 2. Otherwise, the degree of n_j is equal to zero (put "0" into the k column of the degree string); go to Step 3.

Step 9: Check if $j = 0$. If so, stop. Otherwise, go to Step 3.

As mentioned, in the previous section, there may be some issues over the dummy nodes and two-way links. These issues are considered in Step 6.

6.4.3 Evaluation process

Once the chromosome is generated, the next task is to evaluate its fitness. The fitness is measured according to the objective function of Equation 4-13 which is the net total benefit. There are two possible strategies, grid-search for optimal toll and binary toll vector.

Grid-search strategy

In order to evaluate the objective function, the optimal toll level for each cordon (chromosome) must be defined. Since finding the optimal toll level involves solving the optimisation problem in Equation 4-9, the process of evaluating the exact benefit of each chromosome can be very time consuming. Instead, in this method the simple uniform toll regime is assumed in order to ease the evaluation process. This assumption is also consistent with the judgmental design criteria mentioned earlier. Initially, a set of predefined toll levels is defined e.g. 100, 200, and 300 pence. In the experiment in this chapter, eight toll levels are defined which are £0.50, £0.75, £1.00, £1.25, £1.50, £2, £3 and £4. Each chromosome (representing a charging cordon) will be evaluated with each toll level. The toll level producing the highest objective function (net total benefit) will be chosen as the optimal uniform toll for that chromosome which gives the fitness (objective function) of that chromosome.

Toll-vector strategy

An alternative approach is to let GA-AS optimise the uniform toll level simultaneously with the cordon location. With this approach, a binary string representing different toll level is implemented. An initial set of possible toll level should be defined as mentioned earlier, e.g. eight possible toll levels of £0.50, £0.75, £1.00, £1.25, £1.50, £2, £3 and £4. Each cordon will have its own toll string (binary number) and the associated decimal number of the toll string identifies the toll level from the set of possible tolls defined *a priori*. Figure 6-9 shows an example of a toll string (binary number). From this toll string, its associated decimal number is five and hence its toll level is £1.50, which is the fifth toll level from the set defined earlier.

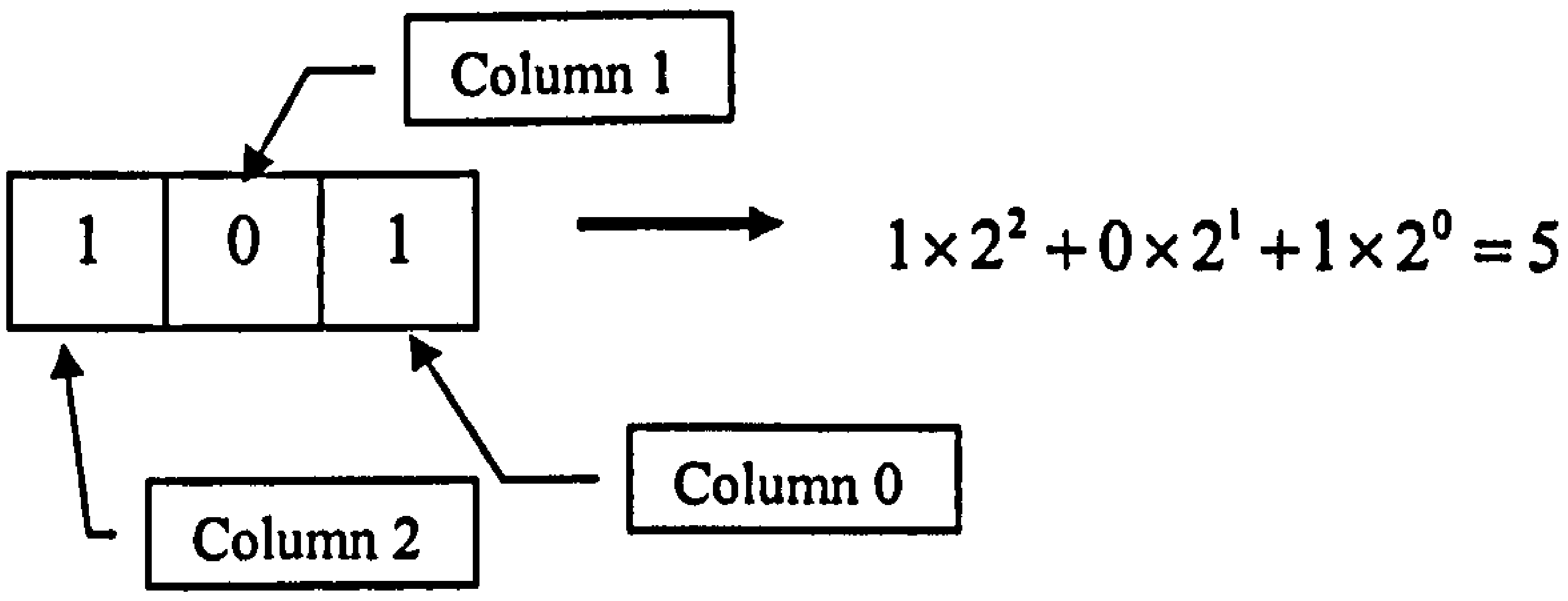


Figure 6-9 Chromosome structure for toll vector

6.4.4 Selection process

The selection process used in GA-AS is based on *'stochastic universal sampling'* which uses a single wheel spin (Michalewicz, 1992). The so called *'roulette wheel'* is constructed where each slot represents a chromosome. In the original form of the roulette wheel, the slots are sized according to the fitness of each chromosome, which represents the probability of a chromosome being selected. However, since the fitness value (total benefits minus scheme costs) in the optimal cordon problem can be negative (which causes a problem for allocating the space for each chromosome on the wheel) the linear ranking approach proposed by Whitley (1989) is adopted. The slots in the roulette wheel are sized according to the chromosome at rank *i*, where the first is the best chromosome, by the following equation:

$$p_i=\frac{1}{|P|}\cdot\left(2-c+(2c-2)\cdot\left(\frac{|P|-i}{|P|-1}\right)\right)$$

where $|P|$ is the size of the population (set *P*), and $1\le c\le 2$ is the *'selection bias'*: higher values of *c* cause the system to focus more on selecting only the better chromosomes. The strongest chromosome in the population can thus be selected with the probability of $\frac{c}{|P|}$; the weakest chromosome can be selected with the probability of $\frac{2-c}{|P|}$. After each chromosome is assigned its probability to be selected, the next step is to calculate a cumulative probability (q_i) for each chromosome:

$$q_i = \sum_{j=1}^{i} p_j$$

Each time, a chromosome is selected for a new population by generating a random number r from the range [0..1]. If $r < q_1$ the first chromosome is selected; otherwise select the i-th chromosome such that $q_{i-1} < r < q_i$. Indeed, some chromosomes would be selected more than once according to the selection probability of each chromosome. As part of the selection process, the idea of *'elitism'* is also adopted to ensure that the best chromosome in the current generation will be included in the population of the next generation.

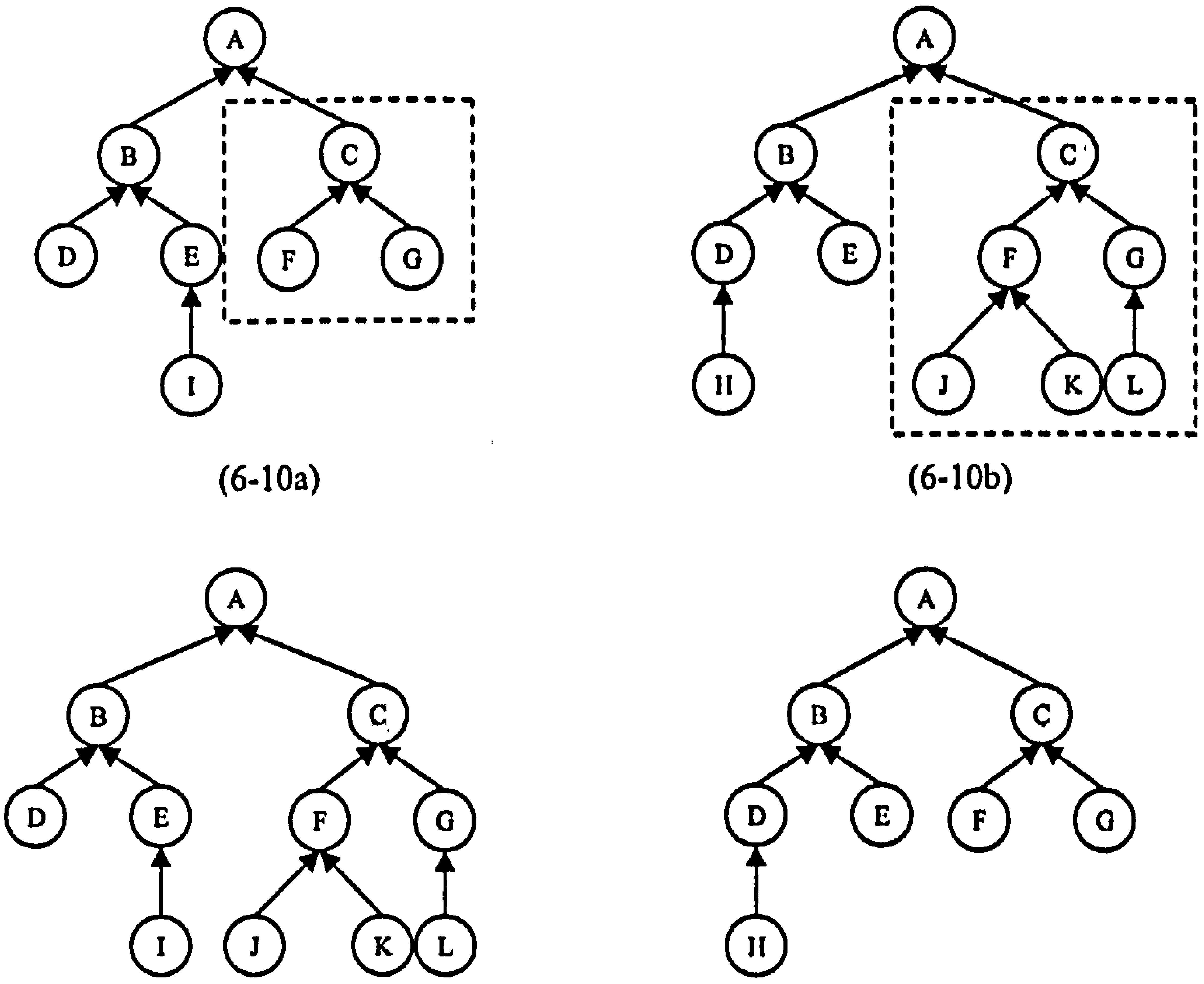
6.4.5 Crossover process

The chromosome structure in GA-AS is in fact very similar to the chromosome used in Genetic Programming (GP), which is also a branch-tree. Thus, it is sensible to adopt the crossover process normally adopted in GP. The crossover process in GP is to cross sub-branches below the chosen nodes in two mated chromosomes. Those who are interested in more details of GP should consult Michalewicz (1992).

The complication involved in crossing the chromosomes in GA-AS is the strict structure of the branch-tree. The process has to start by identifying identical nodes in two mated branch-trees. Then, the crossing node is randomly chosen from the set of identical nodes. The parts of node and degree strings representing the sub-branches in two mated branches rooted from the chosen node are identified. Then, the two sub-branches are crossed over to produce two new chromosomes. Example 6-6 and Figure 6-10 illustrate this process.

Example 6-6 Illustration of crossover process for the branch-trees

From Figure 6-10, there are two original mated branch-trees (in Figure 6-10a and Figure 6-10b). The set of common nodes in these two mated chromosomes (except root node) are nodes B, C, D, E, F, and G. Assume that node C is randomly selected by the crossover process. The sub-branches rooted from node C in two branch-trees are identified which are the parts inside the dash-line boxes in Figure 6-10a and Figure 6-10b. These two sub-branches in the two mated chromosomes are then switched. In other words, the sub-branch in the dash-line box in Figure 6-10a is moved to replace the sub-branch in the dash-line box in the other branch-tree in Figure 6-10b, and vice versa. Figure 6-10c and Figure 6-10d show the two new branch-trees after crossover.



(6-10a)

(6-10b)

(6-10c)

(6-10d)

Figure 6-10 Crossover process for branch-tree

The existence of a dummy node requires the algorithm to check the new chromosome after the crossover operation. After applying GA operators, an algorithm for detecting the new dummy nodes as the result of the crossover operation must be applied to the new chromosomes.

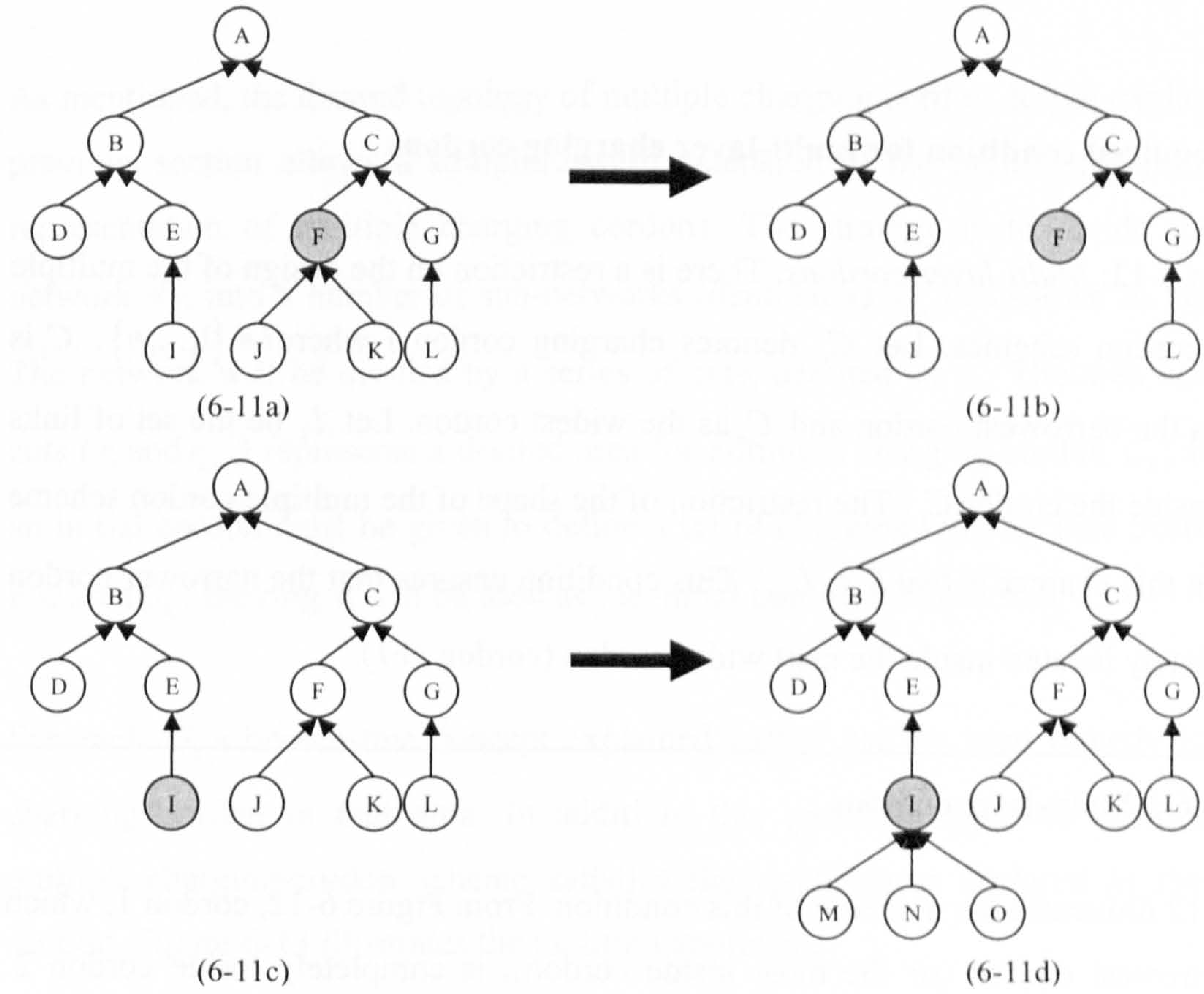
If the toll-vector strategy is adopted to optimise the uniform toll level as well, then the normal crossover process, as discussed in section 5.6.1 in Chapter 5, can be applied to the toll string of two mated chromosomes.

6.4.6 Mutation process

The second GA operator is the mutation process which aims to preserve the diversity amongst the population and to represent the stochastic evolution process in nature. In a typical mutation process, the value of the bit to be mutated in the string will be changed to the opposite value, i.e. if the current value is “1” then it will be changed to “0” and vice versa. However, the chromosome structure in GA-AS is not consistent with the traditional binary string chromosome. Thus, a new approach to mutate the chromosome is developed. The mutation process in GA-AS involves the branching process at a node (including both branching in and out). For a leaf node, if the node is to be mutated, the node will be expanded following the branching process explained earlier (this is the branching out process). Alternatively, if the node to be mutated is not a leaf node, the branch-tree will be branched in at that node, i.e. converting that node to a leaf node and removing all nodes below the mutated node from in a sub-branch. Example 6-7 and Figure 6-11 illustrates how the mutation works.

Example 6-7 Illustration of the mutation process for branch-tree

The branch-tree in Figure 6-11c is used as the original branch-tree in this example. The two branch-trees in Figure 6-11a and Figure 6-11b exhibit the mutation as branching in where node F is contracted. In this example, nodes J and K are removed from the branch-tree as the result of the mutation at node F; the degree of node F is also set to zero. The two branch-trees in the lower part demonstrate the branching out as the mutation process where node I is expanded. Nodes M, N, and O which are the preceding nodes of node I in the full network are added to the branch-tree; the degree of node I is also changed to three.



(6-11a) (6-11b) (6-11c) (6-11d)

Figure 6-11 Mutation process for branch-tree

After the mutation process is applied to a chromosome the chromosome must be checked for any new dummy node. Note that in the branching out process, before adding preceding nodes the two-way link issue must be checked. Similarly to the crossover process, the normal GA mutation method can be applied directly to the toll string if GA-AS is asked to optimise the uniform toll level as well.

6.5 GA-AS FOR MULTIPLE CHARGING CORDON DESIGN

In some cases, a charging cordon scheme may comprise a set of charging cordons as discussed in Chapter 3. By introducing additional cordons, the tolls imposed on each trip can be better adjusted according to its length, contribution to congestion, externalities generated, etc. The GA-AS method proposed in the previous section and the concept of branch-tree framework proposed in Section 6.3 can be extended to deal with the multiple cordon design problem (called MCD problem). This section explains the approach to extend GA-AS to the MCD problem. The extended method is named GA-ASII.

6.5.1 Required condition for multi-layer charging cordons

Definition 6-12: *Multi-layer cordons*. There is a restriction on the design of the multiple charging cordon schemes. Let C_i denotes charging cordon i where $i \in \{1,...,n\}$. C_1 is defined as the narrowest cordon and C_n as the widest cordon. Let L_i be the set of links covered inside the cordon C_i. The restriction of the shape of the multiple cordon scheme adopted in this chapter is that $L_i \subset L_{i+1}$. This condition ensures that the narrower cordon i is completely located inside the next wider cordon (cordon $i+1$).

Example 6-8 Multi-layer cordons

Figure 6-12 illustrates the meaning of this condition. From Figure 6-12, cordon 1, which is the narrowest cordon (or the most inside cordon), is completely inside cordon 2. Similarly, cordon 2 is completely inside cordon 3, which is the widest cordon. This shape of multi-layer cordon scheme facilitates the extension of the branch-tree framework to this problem and it is also consistent with the idea of simple scheme design which enhances the practicality of the scheme.

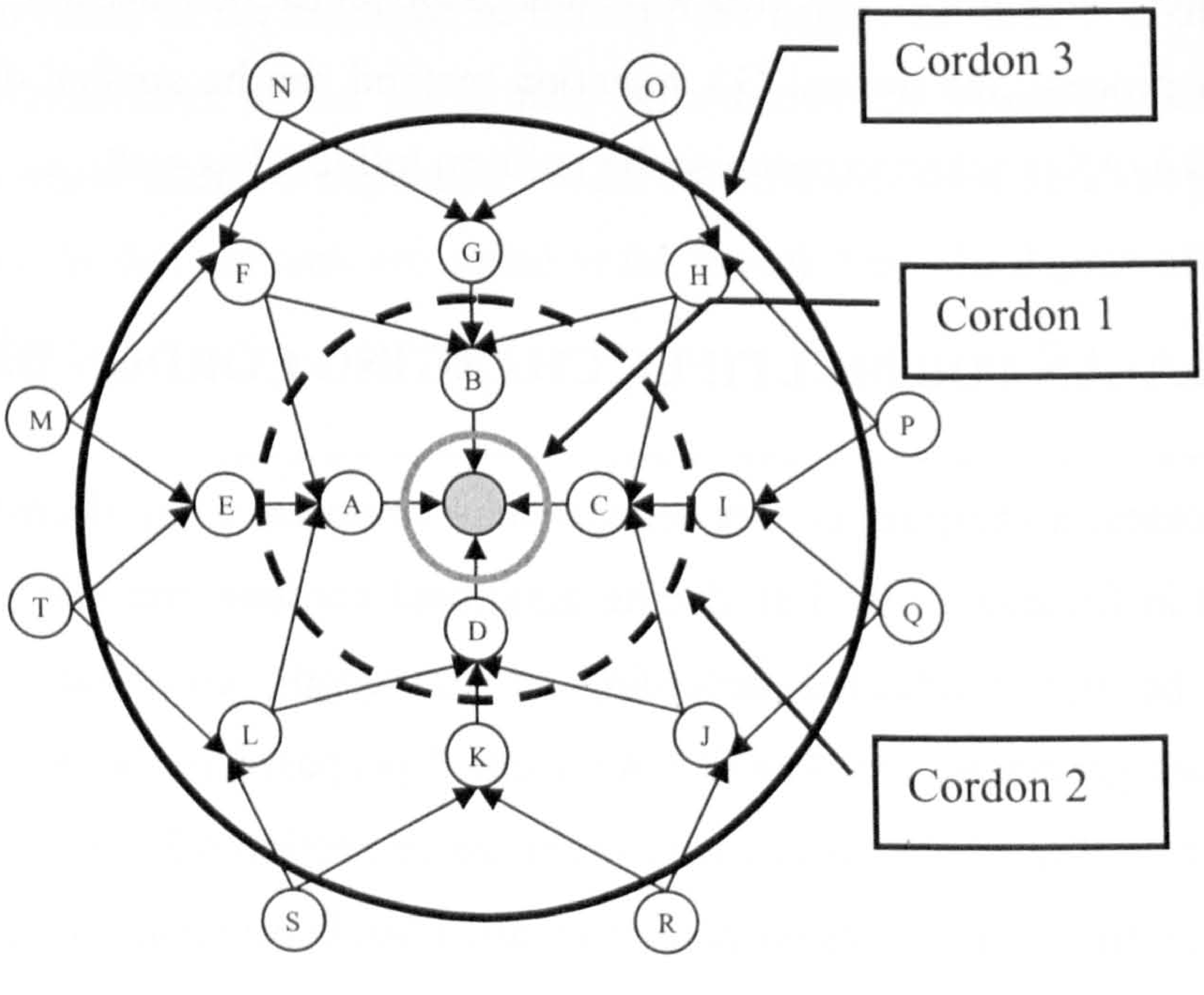


Figure 6-12 Definition of multiple charging cordons scheme

6.5.2 Hyper-graph approach to define multiple charging cordons

As mentioned, the desired topology of multiple charging cordon design explained in the previous section allows a straightforward extension of the branch-tree framework to representation of multiple charging cordons. The strategy is to divide the original network, G, into a number of sub-networks (denoted as G'_k) as shown in Figure 6-13. The network will be divided by a series of cuts, denoted as r_k. The area between two cuts (r_k and r_{k+1}) represents a desired area for putting a charging cordon C_k. Recall that an initial cordon must be given to define a set of charging cordons (see Section 6.3.3). For area G'_k, the ring r_k will be used as the initial cordon for cordon C_k.

For each G'_k a branch-tree concept explained earlier can be used directly to define a charging cordon in that area. In addition, this partition approach ensures that the multiple-charging-cordon scheme satisfies the requirement declared in the previous section. Figure 6-13 illustrates the partition approach.

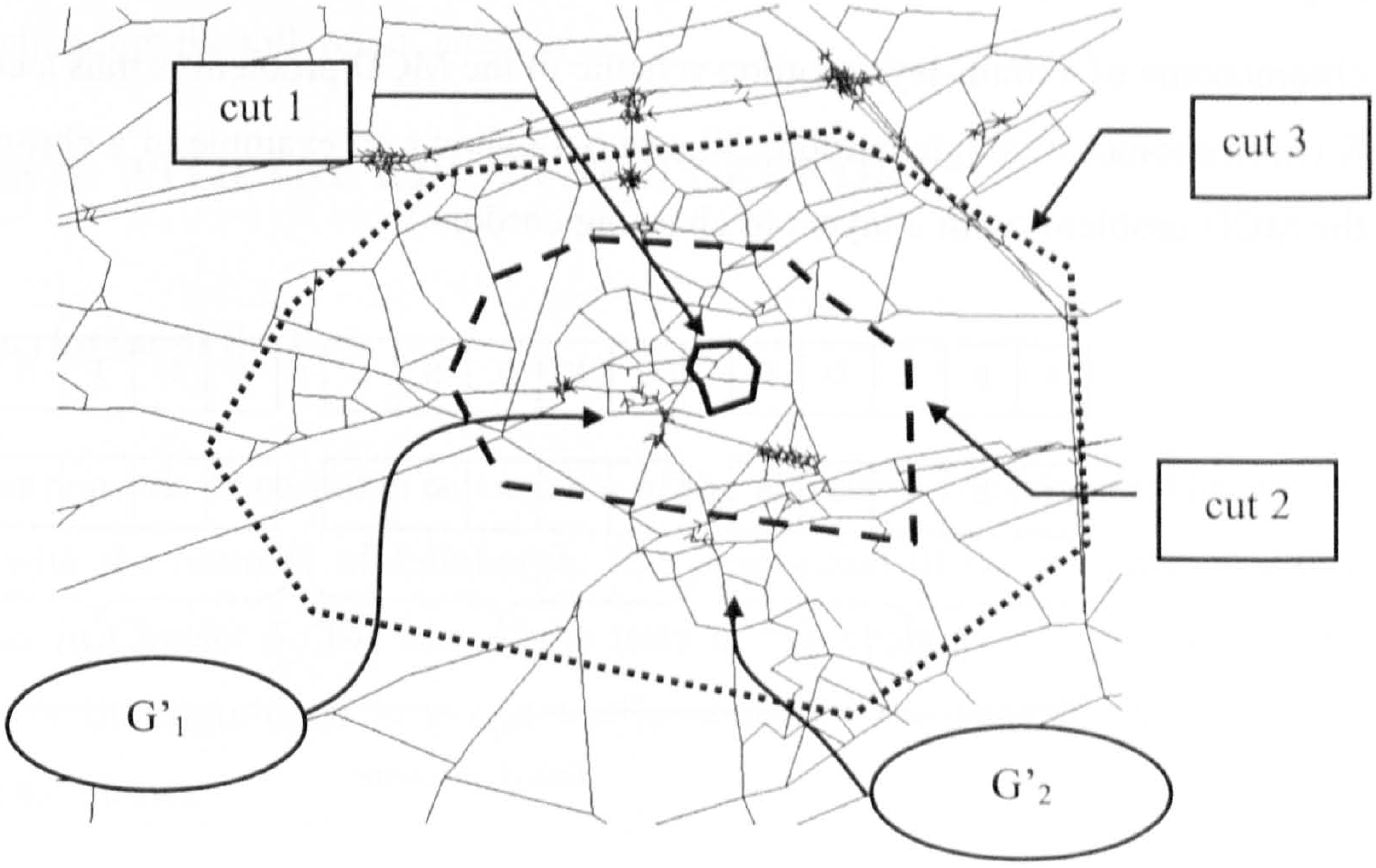


Figure 6-13 Defining multiple charging cordons with cuts

The set of cuts adopted to partition the network are cuts 1, 2, and 3 as shown in the figure. Cuts 1 and 2 define G'_1 which is the area for the first cordon, C_1 (the narrowest cordon). In addition, cut 1 is the initial cordon for the generation of a charging cordon in G'_1.

6.5.3 GA implementation

As mentioned, the requirements of multi-layer cordon and structure of the hyper-graph approach enable a direct implementation of GA-AS and branch-tree approach to the problem. For each G'_k defined, a branch-tree as defined in section 6.3.2 is used to represent a charging cordon C_k. This section explains some necessary modifications of GA-AS to the MCD problem. The key changes required are the chromosome structure and evaluation process.

Chromosome structure

The chromosome structure for a single cordon design is comprised of two strings: node string and degree string (as explained in section 6.4.1). In the MCD problem, two strings are needed for representing each cordon. Therefore, in finding K optimal cordons (C_k for $k = 1 \ldots K$) a number of strings needed for one chromosome is $2 \cdot K$. The chromosome of a multi-layer cordon scheme in the MCD problem is thus a collection of K chromosomes of single cordons. Figure 6-14 shows an example of a chromosome for the MCD problem (with 3 layers of charging cordons).

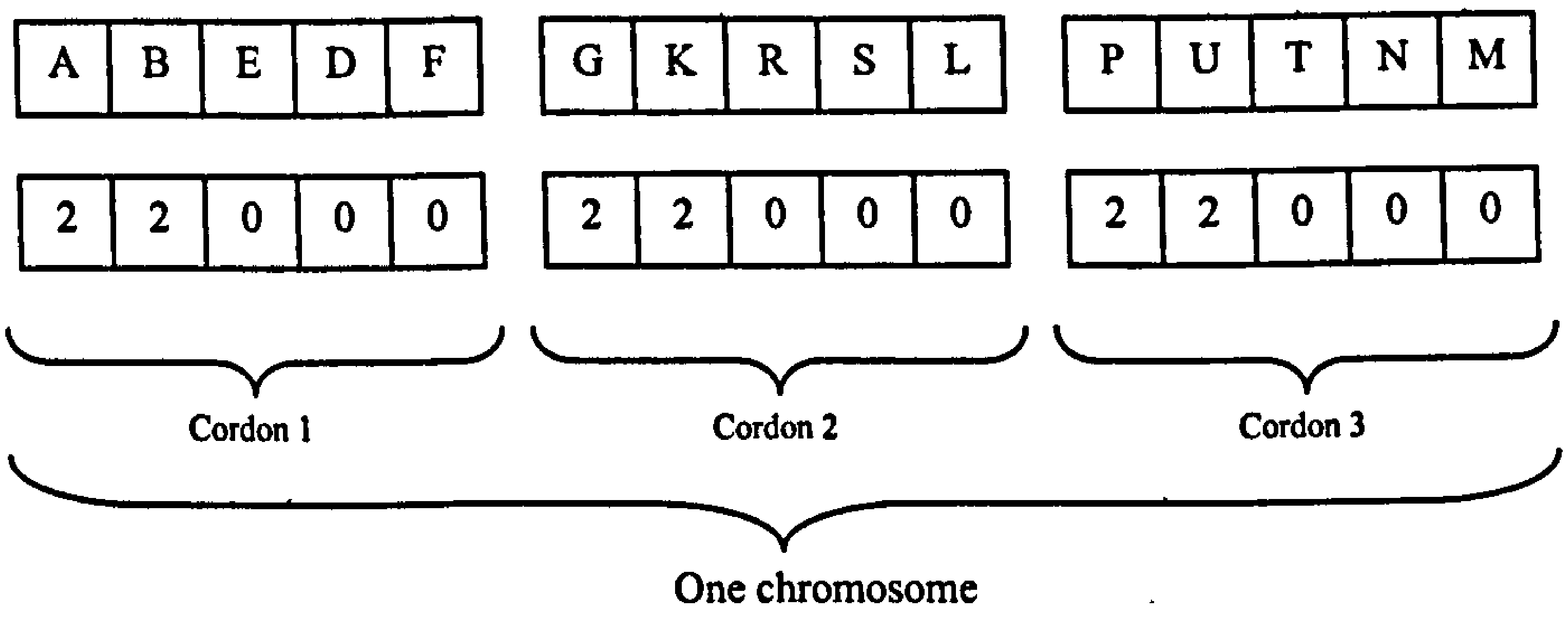


Figure 6-14 Chromosome structure for multiple charging cordons

With the new chromosome structure, the initialisation, crossover, and mutation process explained earlier can be applied to each pair of node and degree strings.

Evaluation

Different toll levels may be allowed for different cordons. One strategy to find an optimal combination of toll levels is to test a multiple cordon scheme with all possible combinations of tolls. This strategy is suitable for the problem with a low number of charging cordons and possible toll levels (e.g. for two cordons with three possible toll levels, the number of possible toll combinations is only nine). However, the number of possible toll combinations can be increased dramatically with the number of cordons and possible toll levels on each cordon. For example, with two cordons and six possible toll levels on each cordon the number of possible toll combinations is 36.

An alternative approach, toll-vector strategy, discussed earlier, which optimises both toll level and cordon location simultaneously, offers a more convenient implementation of GA-AS with the MCD problem. For each solution, a toll string (binary number string) representing the toll level of a cordon is associated with each cordon in that solution. In the case where all cordons will have the same toll level, only one toll string is required. The crossover and mutation process can be applied to these toll strings directly. Given the natural extension of the toll string approach to the MCD problem, GA-ASII will adopt the toll-vector strategy.

6.6 TESTS WITH THE EDINBURGH NETWORK

6.6.1 Description of the tests

This section presents some initial test results of the algorithms developed, GA-AS and GA-ASII, with the network of Edinburgh. The description of the network is already given earlier in Chapter 5. The aim of the tests in this chapter is to demonstrate the application of the algorithms to a real traffic network. The tests conducted in this chapter are as follows:

(i) Use GA-AS with grid-search to find optimal cordon and uniform toll
(ii) Use GA-AS with toll vector to find optimal cordon and uniform toll
(iii) Use GA-AS to find optimal cordon for a given toll level
(iv) Use GA-ASII with and without toll vector to find double cordons with two toll levels.
(v) Test GA-AS with toll vector with different GA parameters, i.e. Pm and Pc

The first, second and third tests aim to demonstrate the methods and show the impact of the constraint on toll level on the benefit of the cordon scheme. The performances of GA-AS with grid-search for toll and toll vector will be compared as well. The fourth test demonstrates the application of GA-ASII in finding the optimal double cordon scheme. The last test is designed to experiment with the effect of the setting of GA parameters. Note that in most cases the objective function being optimised is the net total benefit as explained in Chapter 4. Nevertheless, in order to demonstrate the flexibility of GA-AS in dealing with different objective functions, in the first test GA-AS is also used to define a charging cordon (and toll level) maximising the net revenues (revenues minus costs).

The main intention of the tests in this chapter is to demonstrate the ability of the method developed. Of course, there are many spin-off results which could be discussed from the transport policy perspective. These issues will be discussed extensively later on in Chapter 9.

6.6.2 Comparison between GA-AS with grid-search and toll-vector

Figure 6-15 shows the Edinburgh network with three judgmental cordons. The three judgmental cordons were designed following the judgmental criteria discussed in Chapter 3. The inner cordon 2 is used as the initial cordon for GA-AS and the ring road is used to define the boundary of the charging cordon. The optimal uniform toll level and its associated social welfare benefit were found for each judgmental cordon (see Table 6-2).

GA-AS with grid-search and with toll-vector are then applied to find an optimal cordon and its associated optimal uniform toll. Table 6-1 shows the parameters used for both algorithms (the number of modelling runs are limited to be the same for both methods). Note that for each chromosome GA-AS with grid-search needs eight SATURN runs to find its associated benefit. Figure 6-16 shows the comparison of the plot of the best objective function found so far against the number of modelling runs for GA-AS with the grid-search and toll-vector.

Population number	Generation number	Probability of crossover	Probability of mutation	Number of elitism chromosomes
50	200	0.75	0.15	5

Table 6-1 Parameters used for GA-AS

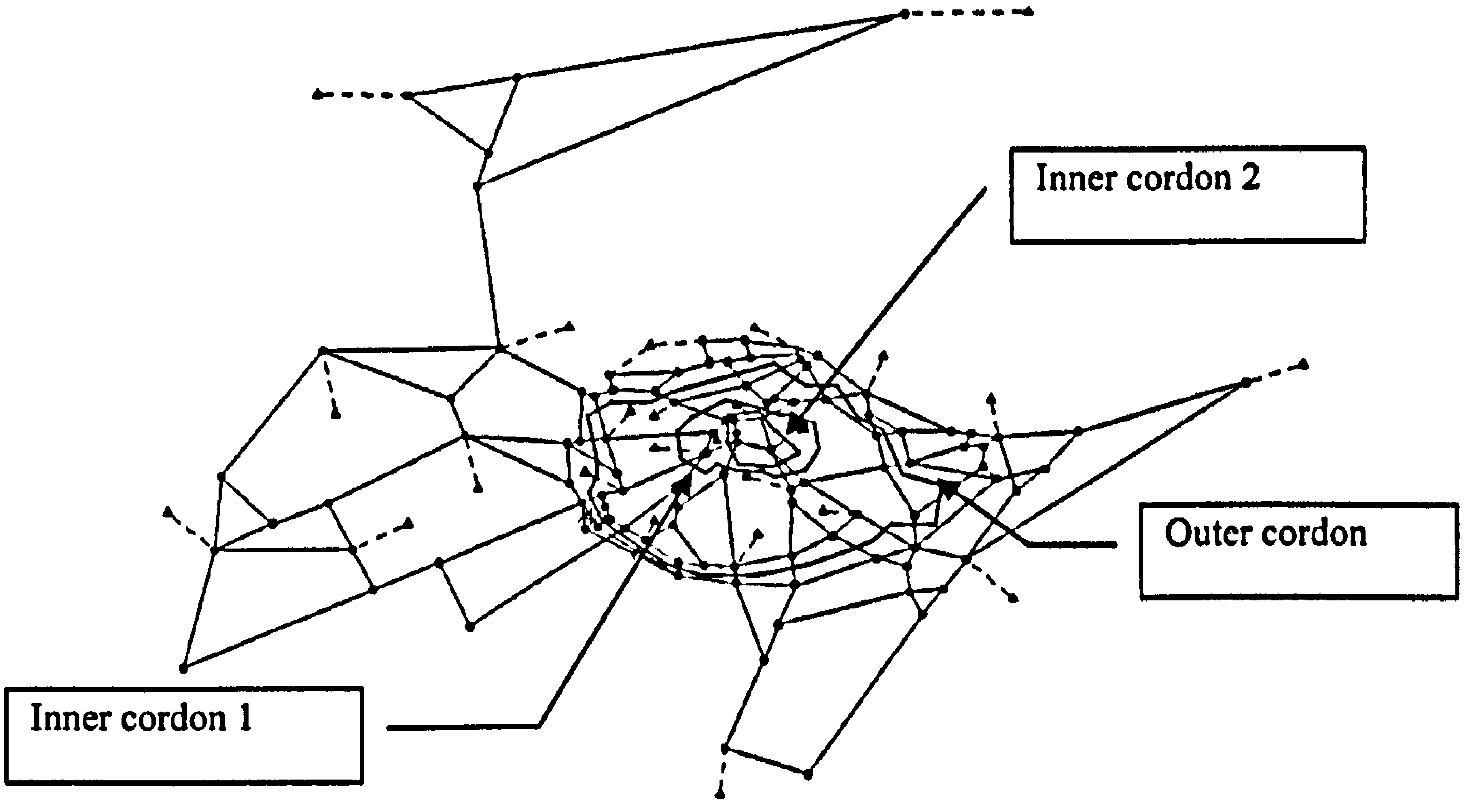


Figure 6-15 Edinburgh network and three judgmental charging cordons

From the results, GA-AS with toll-vector strategy outperformed GA-AS with grid-search strategy. GA-AS with toll-vector found an optimal solution after 4,400 runs whereas GA-AS with grid-search found the same solution after 6,720 runs. With the grid-search strategy, a chromosome with a lower objective function (associated with its optimal toll) is less likely to survive compared to the case with the toll-vector strategy. For example, with the grid-search a chromosome *i* is pooled with *n* other chromosomes.

Assume that other chromosomes (representing different cordons) have substantially higher objective functions with their associated optimal tolls (found from grid-search) compared to the objective function of chromosome *i*. In this case, the probability of chromosome *i* to be selected would be very low. On the other hand, with the toll-vector approach, although chromosome *i* is pooled with the same *n* chromosomes, it may have a higher probability to survive if the toll vectors of *n* other chromosomes produce lower objective functions compared to the toll vector of chromosome *i*.

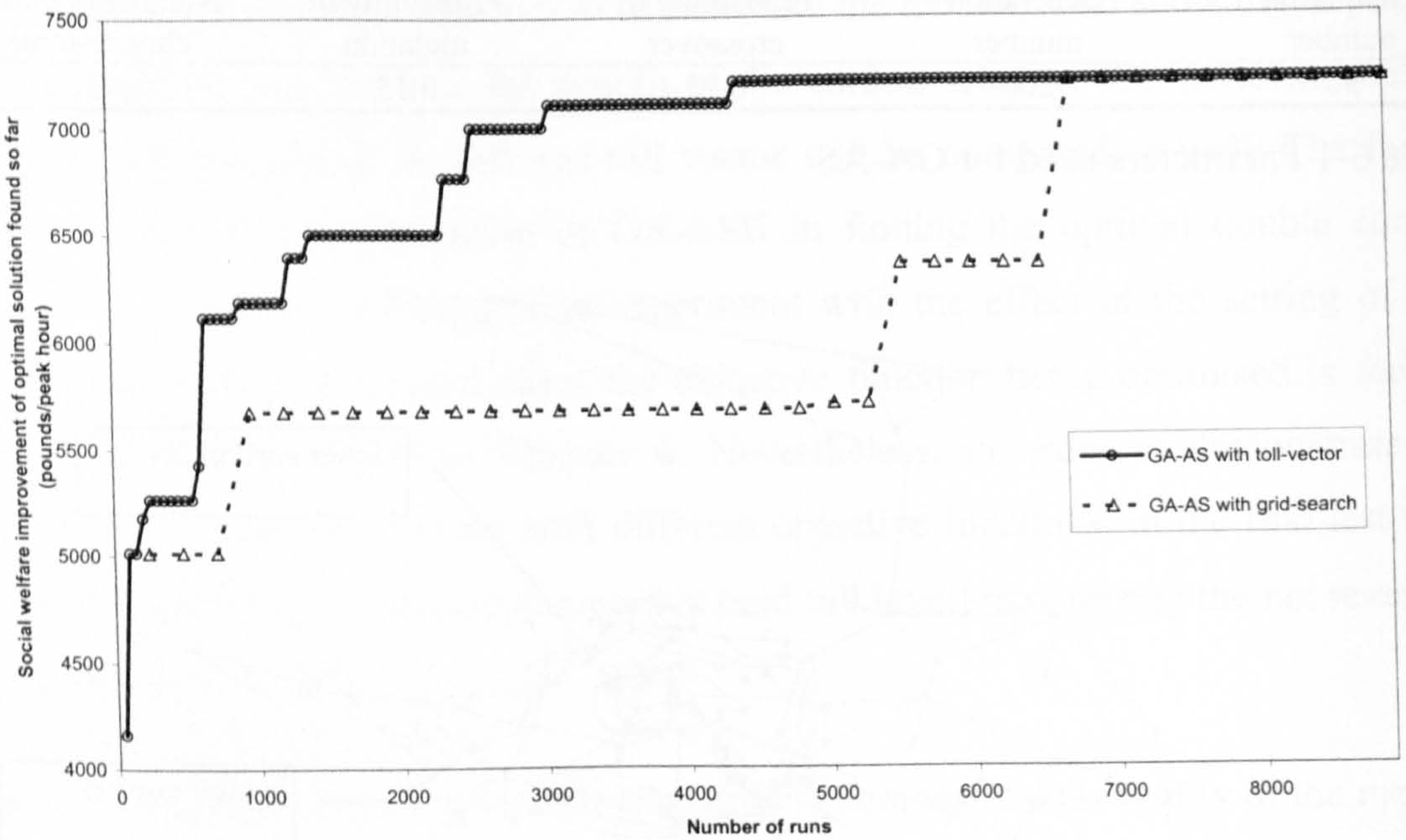


Figure 6-16 Performance comparison between GA-AS with grid-search and toll-vector strategies

The relationship of this phenomenon with the searching mechanism of GA is that GA-AS with the grid-search will have a more limited searching space compared to the searching space of GA-AS with toll-vector. It is well known in theory of GA that some weaker chromosomes may lead the searching algorithm to a better chromosome. Therefore, by associating the fitness of each cordon to its optimal toll the algorithm may lose some weaker cordons whose topology may lead the algorithm to the optimal or better solution. In addition, GA-AS with grid-search may waste too much computational time in applying the grid search without associating the solutions with the crossover and mutation processes which are the search operators for GA.

Figure 6-17 and Figure 6-18 shows the actual distribution of the fitness of the chromosomes from GA-AS with toll-vector and grid-search respectively. The comparison of these two figures demonstrates the advantage of using GA-AS with the toll-vector strategy in terms of the diversity of the solution and searching space. Note that despite a smaller number of plots of the population on Figure 6-18 for GA-AS with the grid-search strategy, the number of simulation runs in both cases is equal.

In fact, this result may not be limited only to the grid-search algorithm. A method for solving a mixed combinatorial optimisation (like optimal toll location and toll level) may perform better if GA is used to simultaneously optimise both variables. For the optimal cordon design problem, it seems to be a better strategy to use GA to simultaneously optimise both cordon location and toll level. However, it should be noted that this analysis is only limited to one case and the results may be vary from case to case.

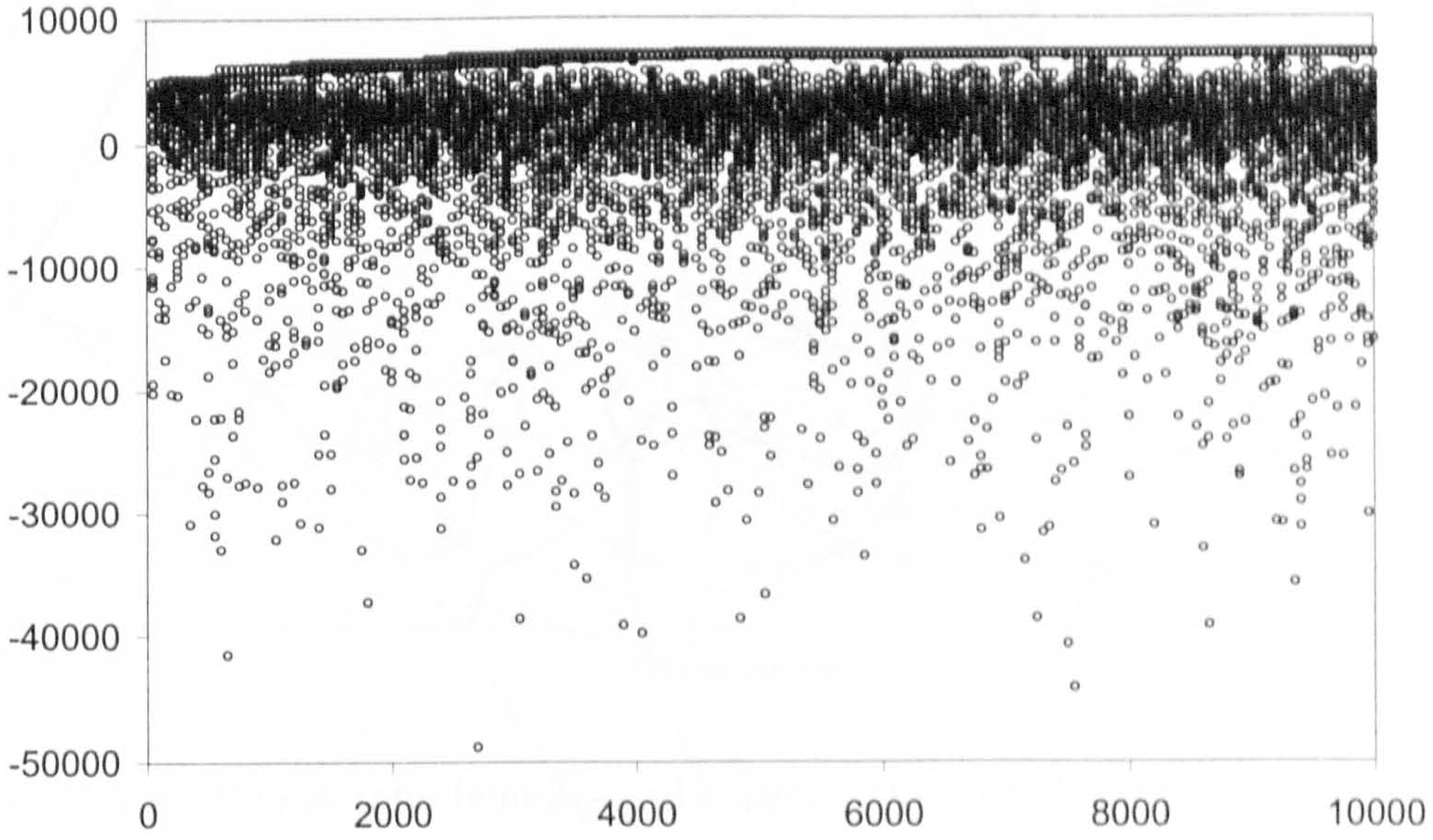


Figure 6-17 Fitness distribution of GA-AS with toll-vector

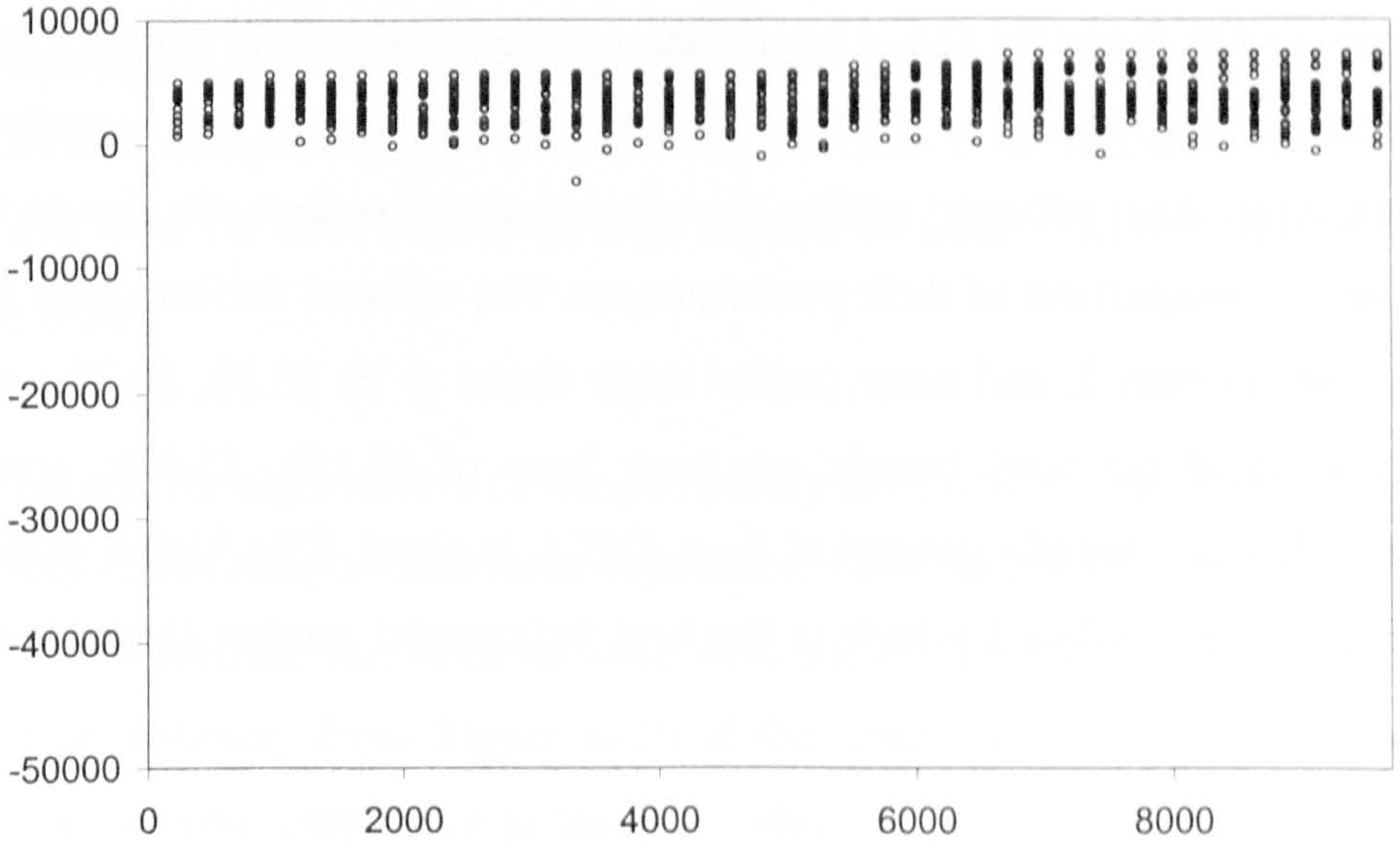


Figure 6-18 Fitness distribution of GA-AS with grid-search

6.6.3 Optimal cordon without a constraint on toll level

Maximise Social Welfare Improvement

Figure 6-19 shows the optimal cordon solution found with the toll level of £1.50. The cordon solution found has 13 tolled links (highlighted links).

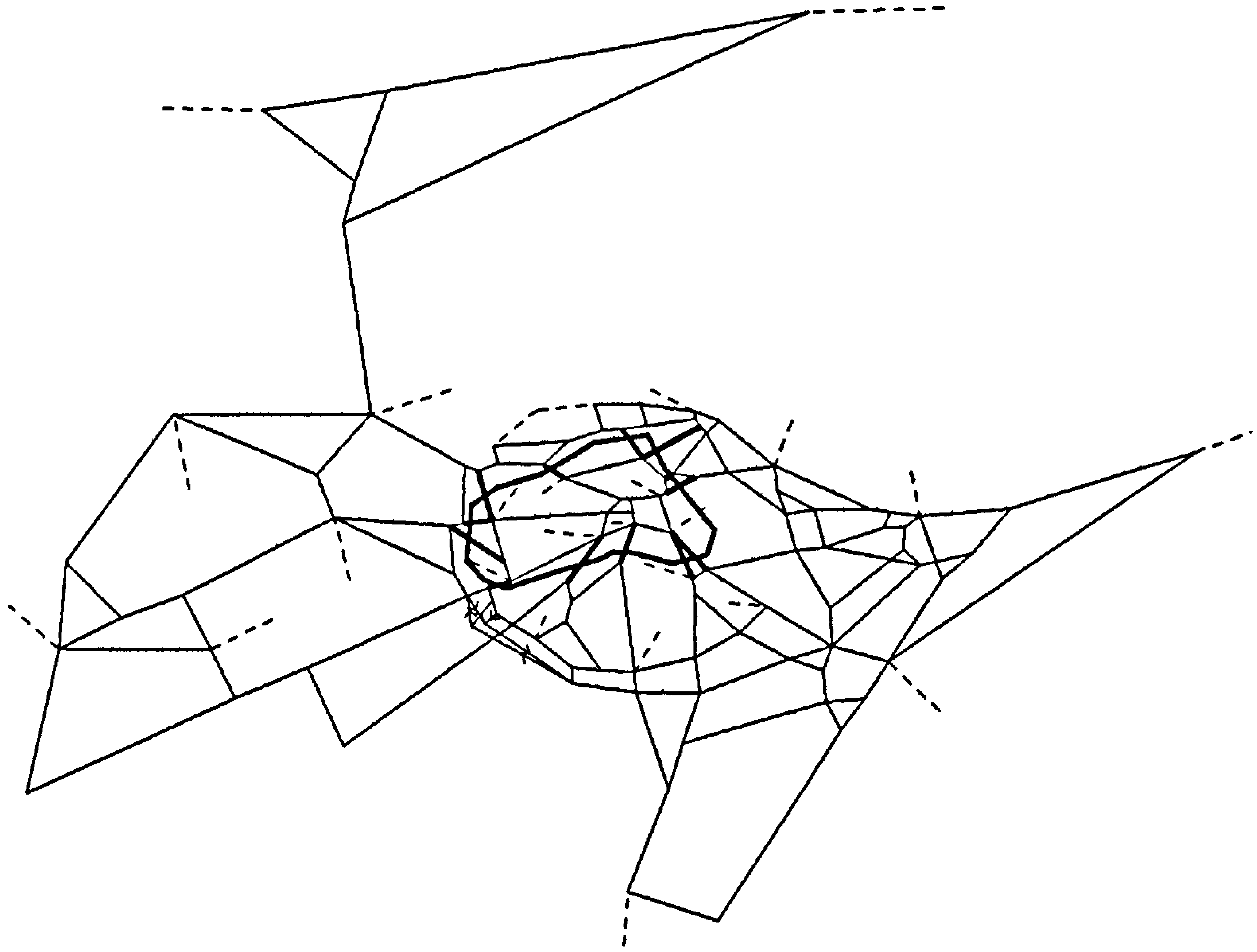


Figure 6-19 Optimal cordon solution found by GA-AS

The optimal cordon found by GA-AS with the toll vector search is named OPC1. The optimal uniform toll found is £1.50. The cordon produces a net benefit around £7.31k/peak hour. As mentioned earlier, three judgmental cordons are also included in the tests for the comparison of their performances. The optimal uniform tolls for inner cordon 1, inner cordon 2, and outer cordon were found to be £0.50, £0.75, and £0.75 with the associated net total benefit per peak hour of £2.10k, £3.99k, and £1.96k respectively. The net benefit generated from OPC1 is about 83% higher than the net benefit of the inner cordon 2 which is the best judgmental cordon (based on the net benefit).

The optimal toll found for OPC1 is £1.50 which is considered a reasonable level of toll. In fact, the current plan of the congestion charging in Edinburgh considers the implementation of £2 double toll rings. It is interesting to see the sensitivity of the net benefit of the scheme when the toll is varied. Figure 6-20 shows the net benefit and revenue curves of the OPC1 with eight different toll levels adopted during the optimisation process.

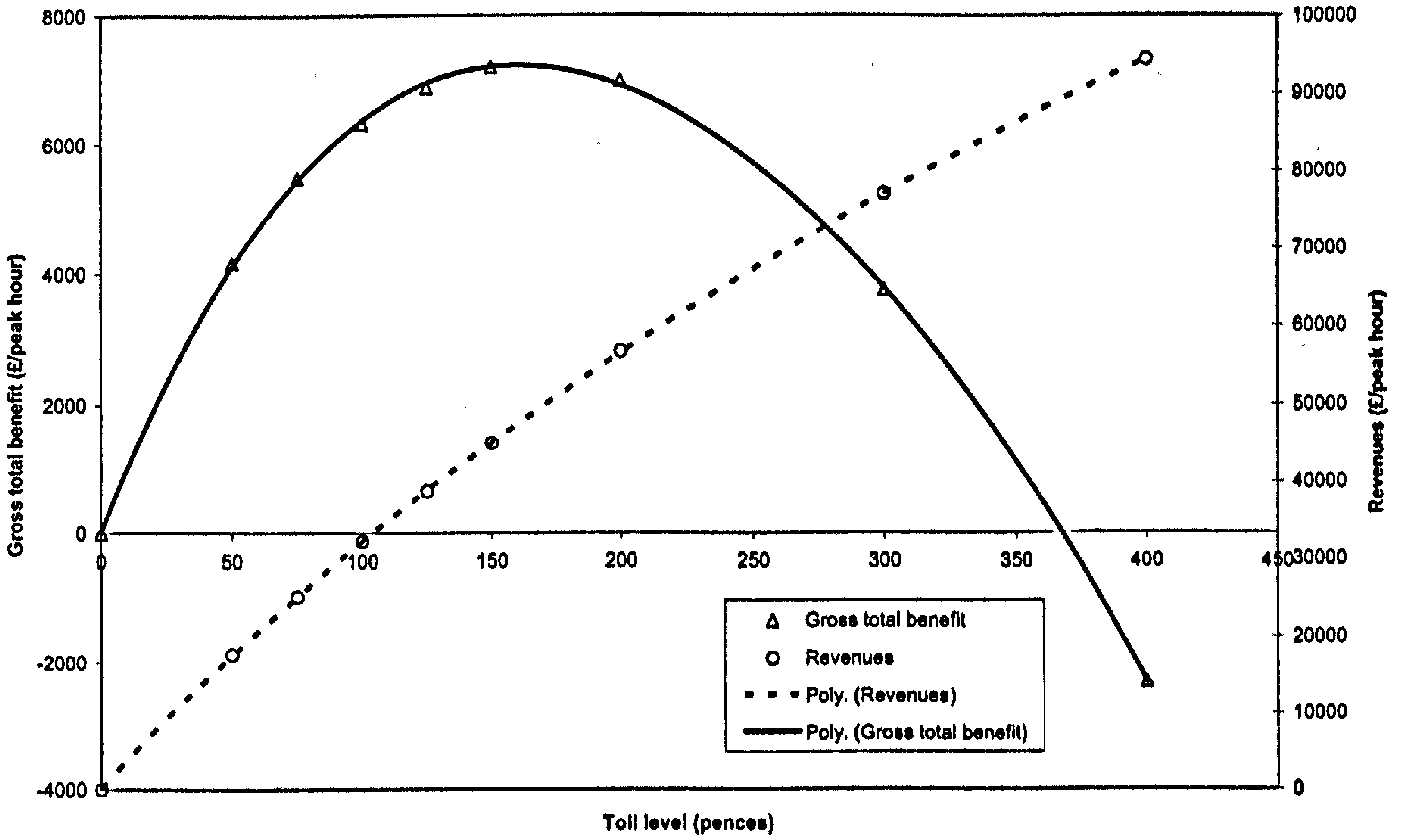


Figure 6-20 Variation of total benefits and revenues for OPC1 cordon with different toll levels

Figure 6-20 confirms the optimal toll level found by GA-AS with toll vector (approximately £1.50) and shows that the benefit of the cordon varies significantly with the toll level. Unsurprisingly, if the objective function is to maximise the revenues, the optimal toll level for OPC1 could be well beyond £4 observing from the revenue curve in Figure 6-20.

In this Chapter, the algorithm is developed purely to optimise a single objective function without any constraints or additional objectives. However, the framework set earlier in Chapter 6 emphasises the need to take account of design constraints and multiple criteria. For instance, from Figure 6-20 if the constraint on the minimum level of revenues is imposed (say to be at least £70,000 per peak hour), then the optimal toll level of £1.50 as found by GA-AS is no longer a feasible solution (the revenue generated from £1.50 toll is only £45,100). Instead, the toll level of £3 would be chosen as the best toll level satisfying the constraint on the guaranteed level of revenues. Similarly, the constraint over issues like equity, congestion relief, or increase in travel distance could also be incorporated into the charging cordon design. These issues will be addressed separately in Chapter 7.

Apart from the best cordon found so far (OPC1), during the GA-AS optimisation process we also can identify the cordons producing the second and third highest net total benefit (named OPC2 and OPC3 respectively). Figure 6-21 shows the location of OPC2 and OPTC and Table 6-2 shows the net total benefit and optimal toll level of these two cordons.

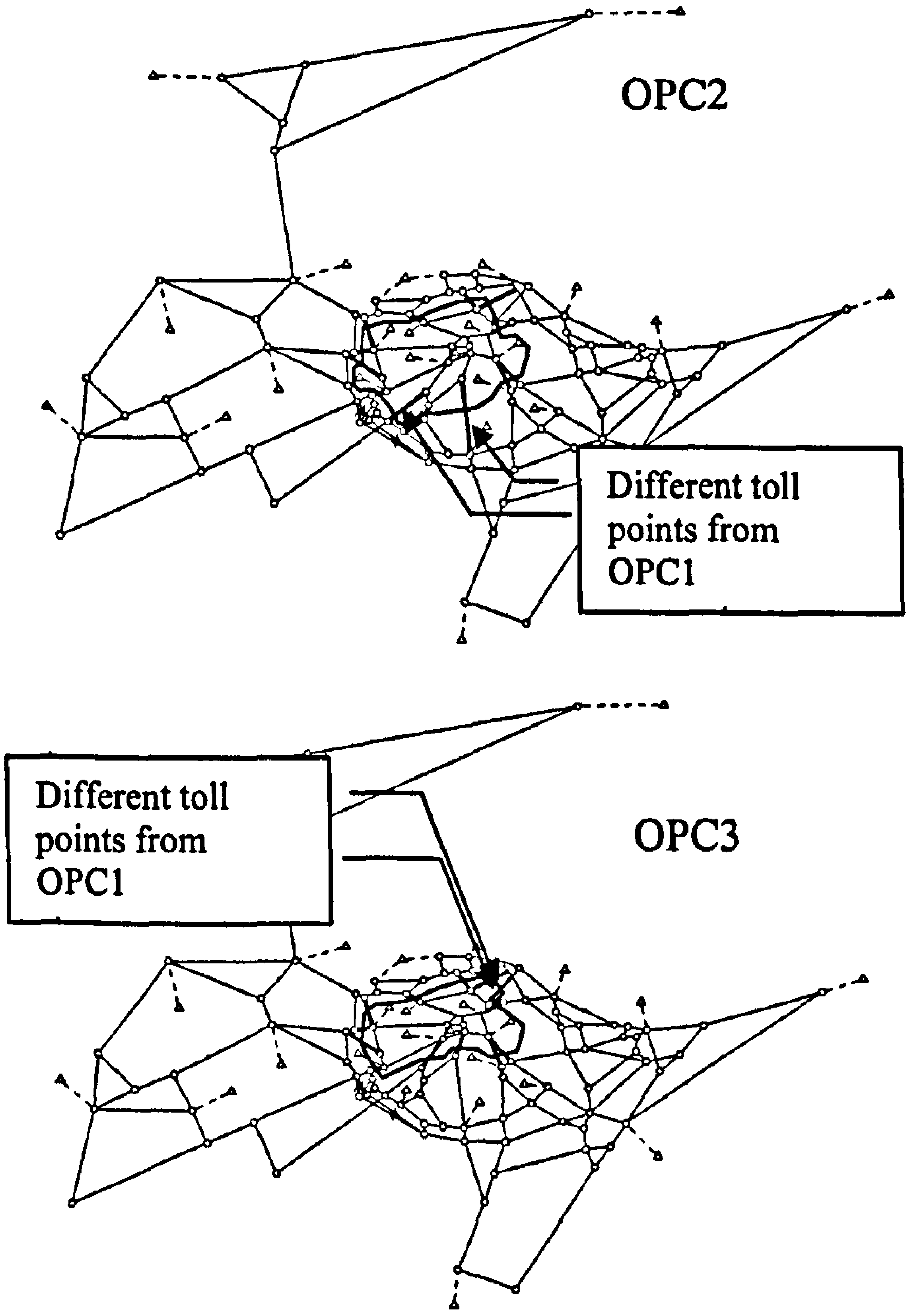


Figure 6-21 Cordons with 2nd and 3rd highest net total benefits

OPC2 and OPC3 have one more toll point compared to those of OPC1. There are two different locations of tolled links from those of both OPC2 and OPC3 compared to the tolled links of OPC1. These links for OPC2 and OPC3 are pointed out in Figure 6-21. Interestingly, despite two different locations of toll points, the total benefits of OPC1, OPC2, and OPC3 are almost identical (around £8.51k/peak hour). Thus, the higher net benefit of OPC1 is mainly due to a smaller number of toll points. It is interesting that a charging cordon with a smaller number of toll points (like OPC1) is capable of generating the same or even higher benefit compared to OPC2 and OPC3. This re-confirms the importance of careful selection of the location of charging cordons and toll points.

Maximise Net Revenues

The objective being optimised in the previous subsection is the net total benefit. In this section, GA-AS is used to optimise the net revenues instead (using the same GA parameters as shown in Table 6-1). Figure 6-22 shows the best cordon and uniform toll level found for maximising the revenues. The optimal toll level for this charging cordon is £4, as expected. This charging cordon is named OPC-REV. OPC-REV consists of 32 toll points which is substantially higher than the number of toll points of OPC1. Table 6-2 shows the main result of the test with OPC-REV. From the result, OPC-REV generates approximately £111.1k /peak hour. On the other hand, it also causes substantial negative impact in terms of the social welfare (with £-41.19k / peak hour of social welfare disbenefit). A combination of the high number of charging point, high toll level, and a wide range of the coverage of the routes in the network enables OPC-REV to generate such a high revenues.

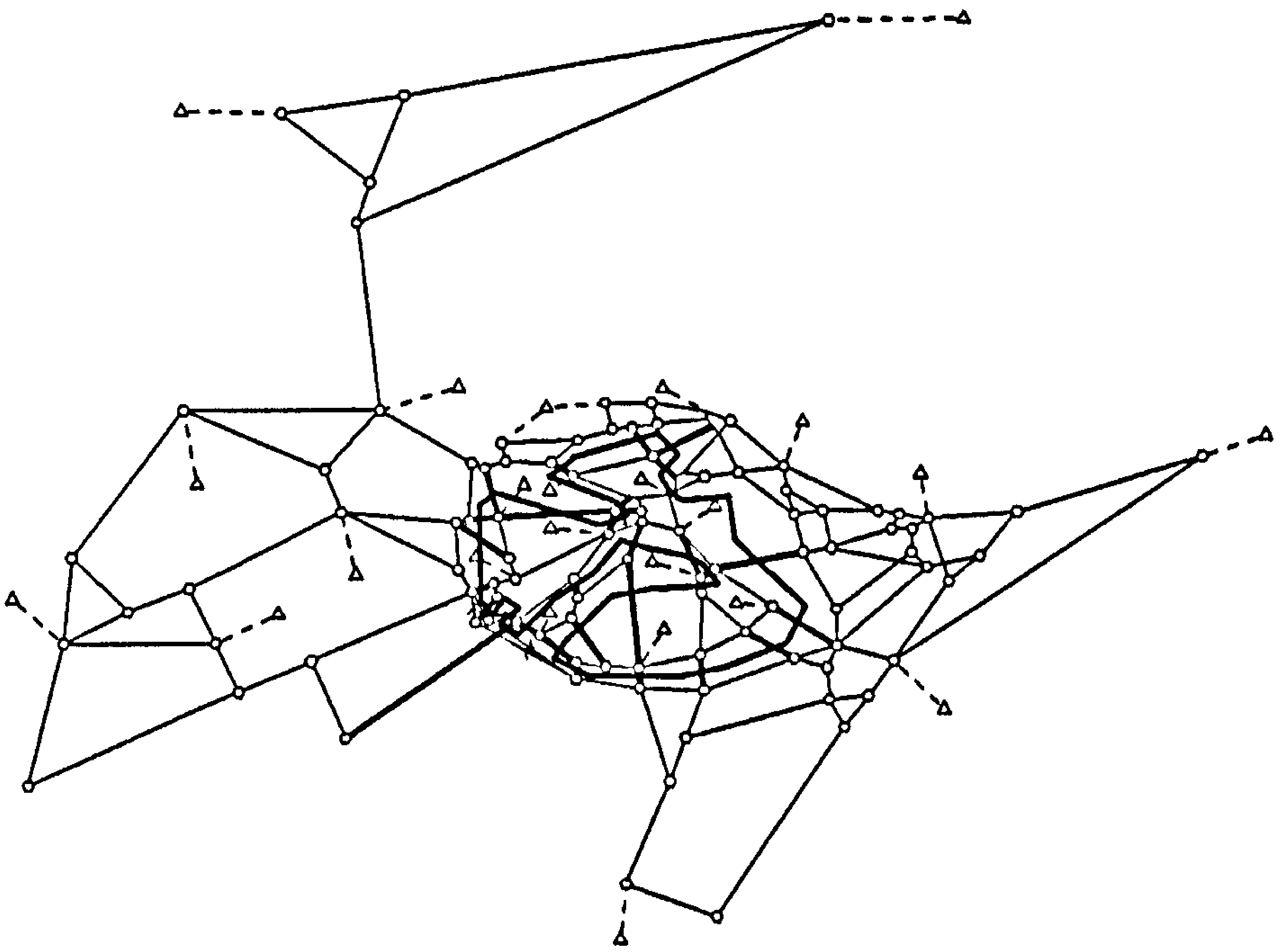


Figure 6-22 Location of OPC-REV (maximise net revenues)

6.6.4 Optimal cordon for each toll level

From the survey results presented earlier in Chapter 3, in some cases the local authorities already have their preferred/feasible toll levels in mind and they only need to find out the best location for the charging cordon. In this circumstance, GA-AS can be used to optimise the location of charging cordon given a fixed toll level. This section presents the results from optimising the locations of charging cordons for different eight toll levels of £0.50, £0.75, £1.00, £1.25, £1.50, £2.00, £3.00, and £4.00.

Figure 6-23 shows the benefits from eight cordons. Figure 6-24 shows the optimal cordon found for the toll levels of £0.50, £0.75, £1.00, £1.25, £2.00, and £3.00. For the toll level of £4.00, the best cordon found cannot generate any benefit. This implies that the toll level aiming to maximise social welfare for this network should not exceed £4. GA-AS also found the same optimal charging cordon for the toll level of £1.50 as tested in section 6.6.3 (see Figure 6-19). The number of tolled points for the cordons in 8-24 (a)-(f) is 12, 10, 10, 13, 11, and 11 respectively. The performances of these cordons (referred to as OPC-τ where τ is the toll level) are presented in Table 6-2.

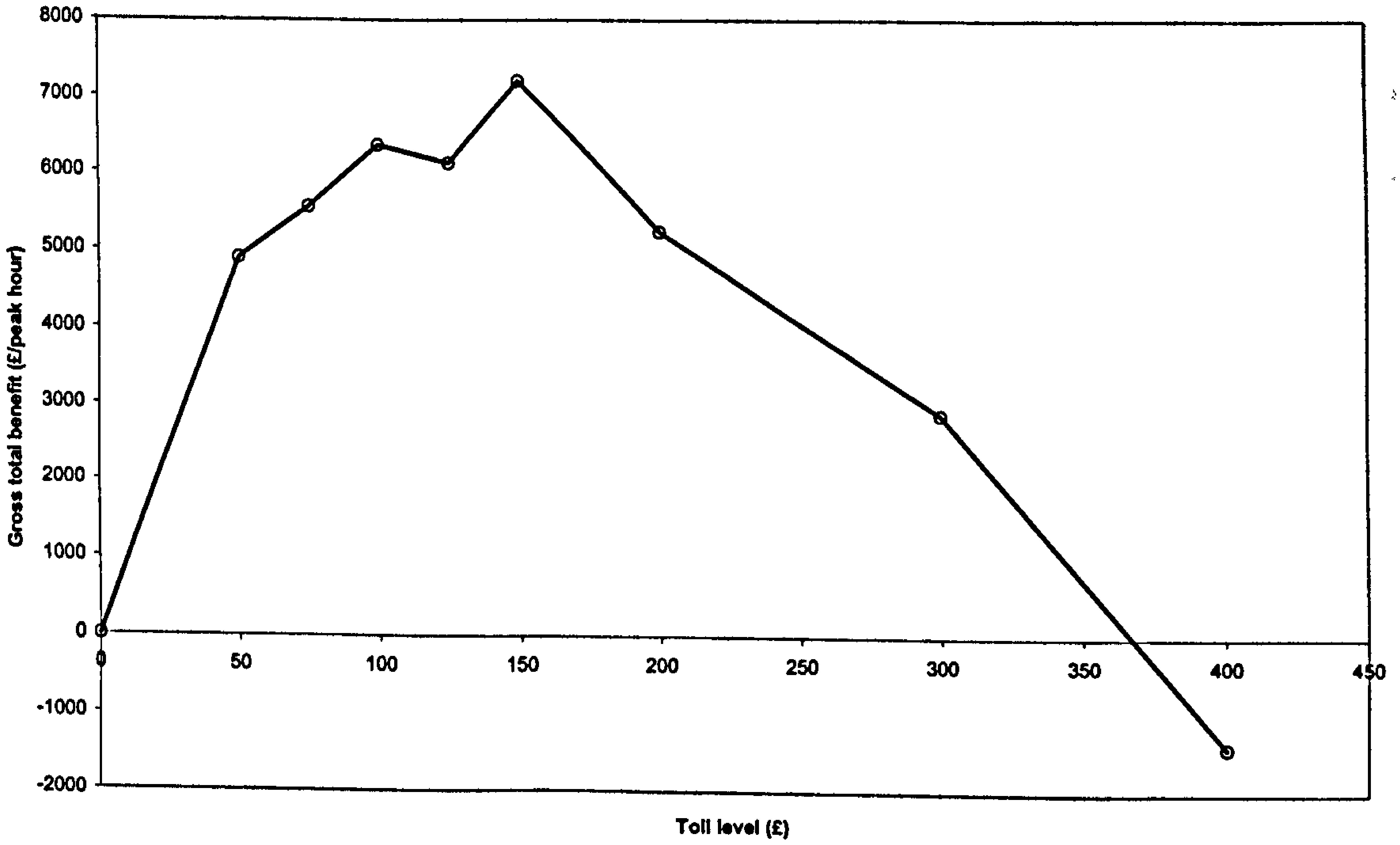
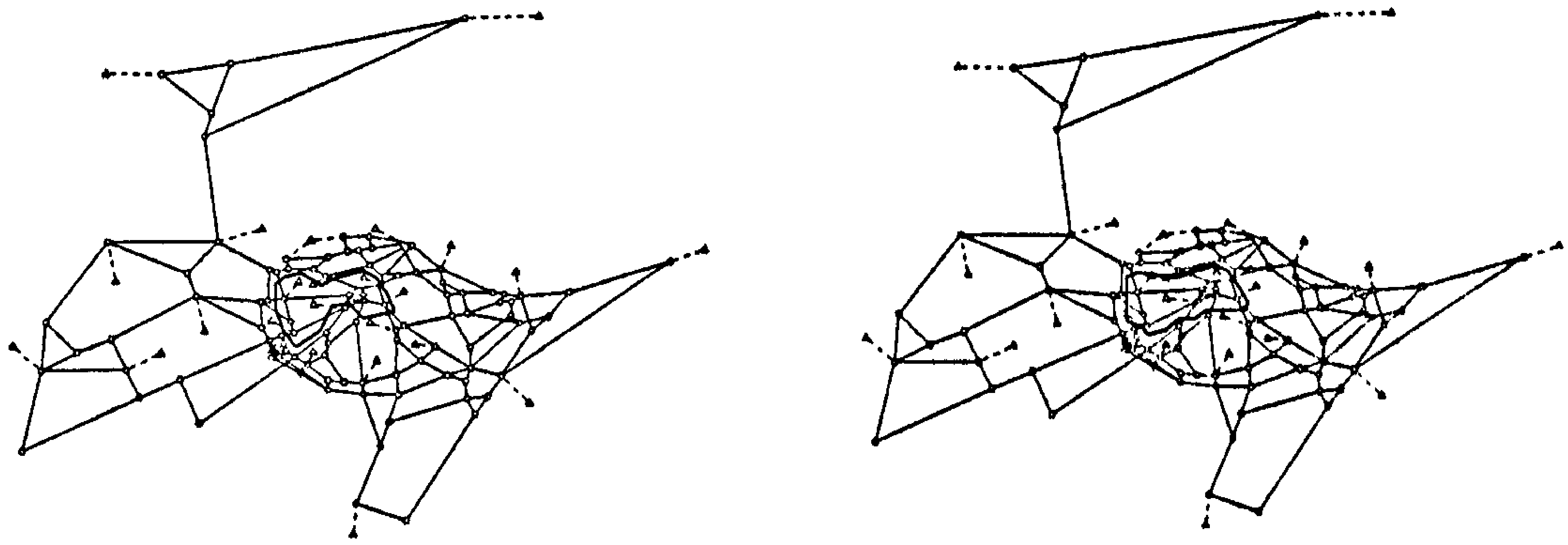
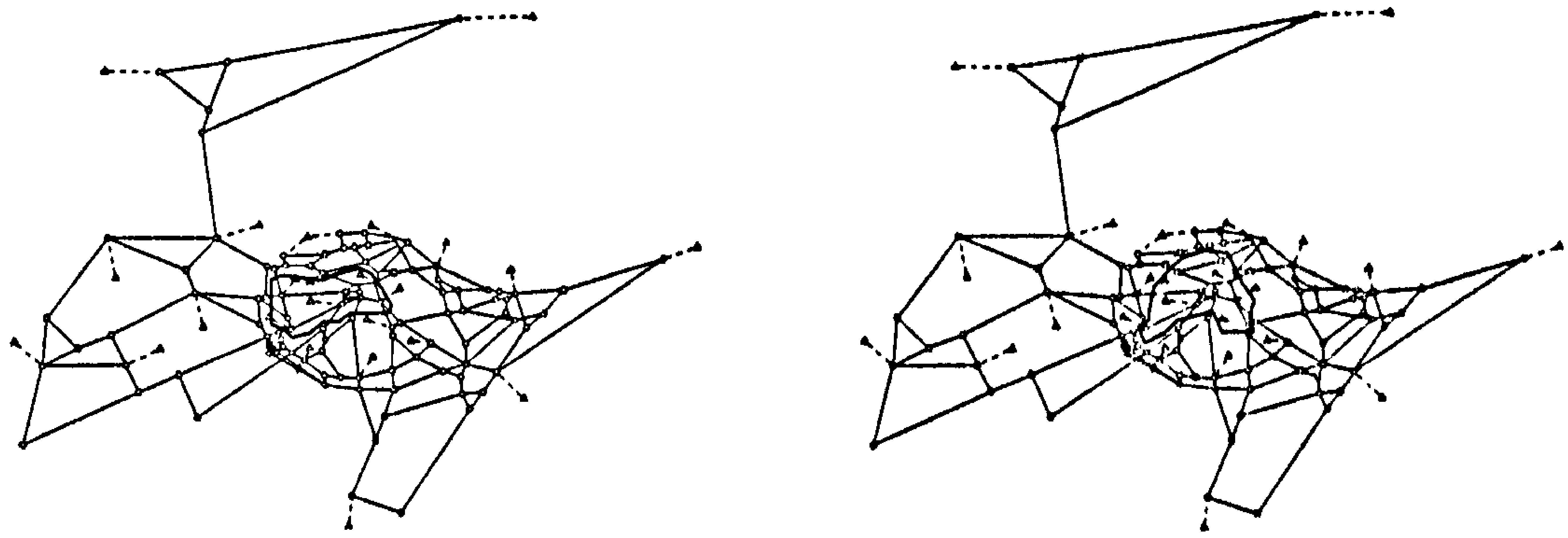


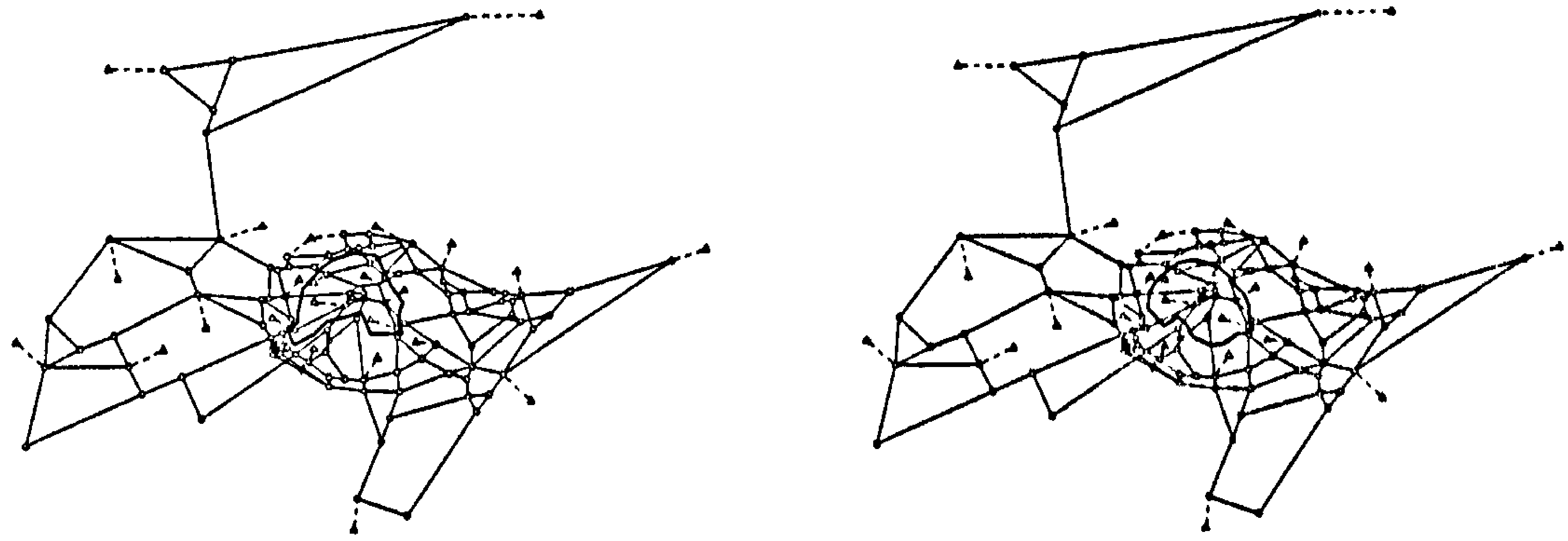
Figure 6-23 Variation of the benefit from optimal charging cordon for different fixed toll levels



(a) optimal cordon for toll level of £0.50 (b) optimal cordon for toll level of £0.75



(c) optimal cordon for toll level of £1.00 (d) optimal cordon for toll level of £1.25



(e) optimal cordon for toll level of £2.00 (f) optimal cordon for toll level of £3.00

Figure 6-24 Locations of optimal cordons for toll levels of £0.50, £0.75, £1.25, £2.00, £3.00

6.6.5 Multiple cordons design

In Section 6.5, GA-AS and the branch-tree framework was extended to deal with the design of optimal multiple charging cordons (referred as the MCD problem). The new algorithm is named GA-ASII. In this section, GA-ASII was adopted to find optimal double cordons with two toll levels (one on each cordon). GA-ASII is applied with the toll-vector strategy for optimising double toll levels. The parameters adopted for GA-ASII in the test are the same as presented earlier in Table 6-1.

Figure 6-25 shows the locations of optimal double charging cordons found by GA-ASII. Although the toll levels were allowed to vary between the two cordons, the optimal uniform tolls found for the inner and outer cordons are both £1.25. The net benefit from this double cordon (D-OPC) is around £15.3k per peak hour which is more than double the net benefit of OPC1. The inner and outer cordons are comprised of nine and 29 toll points respectively giving the total number of toll points of 38. The revenue generated from D-OPC is about 5% lower than the revenue of OPC-REV. Figure 6-26 shows the surface of the net total benefit of D-OPC when the toll levels on the inner cordon and outer cordon are varied. The optimal tolls for inner and outer cordons from the surface are both £1.25. Differentiating the tolls on two cordons did not bring in additional benefit. Note that this result is only specific to this network and may not be true in other cases.

Figure 6-25 Optimal double charging cordons for the Edinburgh network

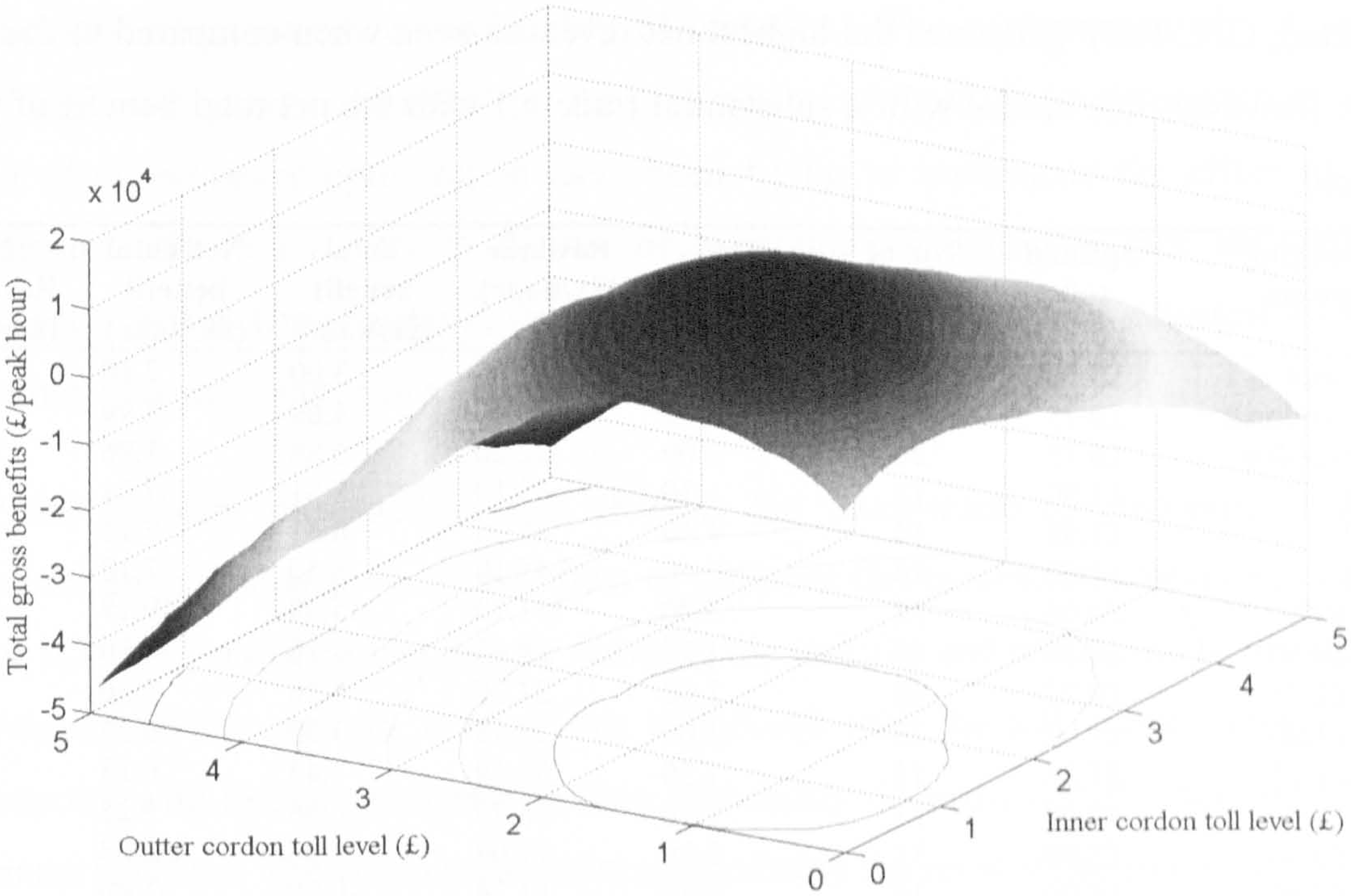


Figure 6-26 Variation of the net benefit of the optimal double cordon with different toll levels

6.6.6 Overall comparison

This section summarises the results from all the tests explained earlier. Table 6-2 presents the key performance indicators for all the tests conducted. As explained, there are five different sets of results. The first group is the test with the judgmental design (inner cordon 1, inner cordon 2, and outer cordon). The second test is the optimal single cordon design maximising the objective function of net total benefit (OPC1, OPC2, and OPC3). The third is the optimal single cordon design but with the objective of maximising the net revenue (OPC-REV). The fourth set of tests is finding an optimal cordon for each given uniform toll level (OPC-£0.50, OPC-£0.75, OPC-£1.00, OPC-£1.25, OPC-£2.00, OPC-£3.00, and OPC-£4.00). The final test is to find the optimal double charging cordon maximising the net total benefit (D-OPC).

As explained earlier in 6.6.3, the cordons designed by applying GA-AS outperform the judgmental design. The improvement of the benefit from the careful design is approximately 80% better than the performance of the best judgmental design. The optimal toll level for the optimised design is around £1.50 whereas the optimal uniform toll for the judgmental cordons is around £0.50 - £0.75. In terms of the net revenues, as expected, OPC-REV generates the highest net revenues even when compared to the D-OPC. However, this comes with a substantial trade-off with the net total benefit of the scheme.

Charging system	Optimal toll	No. of toll points	Cost (£k/hour)	Revenue (£k/hour)	Total benefit (£k/hour)	Net Total benefit (£k/hour)	Net Revenues (£k/hour)
Inner cordon 1	£0.50	9	0.90	12.60	3.00	2.10	11.70
Inner cordon 2	£0.75	7	0.70	18.40	4.69	3.99	3.99
Outer cordon	£0.75	20	2.00	22.20	3.96	1.96	1.96
OPC1	£1.50	13	1.30	45.10	8.51	7.21	43.70
OPC2	£1.50	14	1.40	45.00	8.51	7.11	43.60
OPC3	£1.50	14	1.40	45.10	8.50	7.10	43.70
OPC-REV	£4.00	32	3.20	111.10	-45.99	-49.19	107.90
OPC-£0.50	£0.50	12	1.20	19.47	6.10	4.90	18.27
OPC-£0.75	£0.75	10	1.00	24.79	6.56	5.56	23.79
OPC-£1.00	£1.00	10	1.00	25.15	7.35	6.35	24.15
OPC-£1.25	£1.25	13	1.30	38.59	7.43	6.13	37.29
OPC-£2.00	£2.00	11	1.10	48.77	6.35	5.25	47.67
OPC-£3.00	£3.00	11	1.10	77.04	3.98	2.88	75.94
OPC-£4.00	£4.00	12	1.20	81.24	-0.20	-1.40	80.04
D-OPC	£1.25, £1.25	38	3.80	105.30	19.08	15.28	101.50

Table 6-2 Summary of the results of designing a single optimal charging cordon for Edinburgh network

The performance of the cordon design also varies significantly by the toll level. With the toll level of £1.50, the charging cordon produces the highest net total benefit (OPC1). On the other hand, when the toll level is only limited to £1.00, the optimised cordon can generate a benefit of around 88% of the benefit generated by OPC1. The optimised design with the toll level of £3 can only generate a benefit of around 40% of the highest benefit achieved. If the toll level is increased to £4, the optimised design cannot generate any benefit. On the other hand, the designs with a high toll level (£3 and £4) generate substantial revenues compared to the designs with lower tolls. This is consistent with the design of OPC-REV where the toll level of £4 is found to be the optimal toll for maximising the net revenue. Introducing the second cordon, D-OPC, obviously increases the performance of the charging cordon scheme. The benefit generated by D-OPC is around 110% higher than the benefit from OPC1.

6.6.7 Analysing the effect of GA parameters

For a meta-heuristic search algorithm like GA, there are various parameters the user needs to predefine (for GA these include generation number, population number, probability of crossover, or probability of mutation). The performance of GA search may depend largely on the setting of the parameters involved. Several researchers in the field of evolutionary optimisation have been trying to investigate the effect of each parameter and find a way to define these parameters optimally (see for example Goldberg, 2002). Unfortunately, the most advance result on optimal adjustment of GA parameters is still limited to a very simple problem.

For the problem presented in this chapter, the tailor-made chromosome structure obscures analytical analysis of the effect of different GA parameters. Nevertheless, this issue is still a very important one for ensuring the stability and robustness of the method developed. In this section, several tests of GA-AS with the toll vector strategy were conducted with two parameters including probability of crossover (Pc) and probability of mutation (Pm). These two parameters are probably the most influential in the success of the GA application since they are the main parameters governing the searching mechanism of GA. The base setting of other parameters including population number and generation number is set following the values shown in Table 6-1.

Five different values of the Pc and four different values of Pm are tested. For Pc these were 0.15, 0.35, 0.55, 0.75, and 0.95 and for Pm these were 0.05, 0.15, 0.25, and 0.35. Figure 6-27 and Figure 6-28 show the comparison of the performance of the GA-AS (with toll vector strategy) with different values of Pc and Pm respectively. The indicator of the performance adopted is the highest net total benefit achieved in each test (with fixed populations and generation numbers). The tests with the Pc of 0.15, 0.35, and 0.75 produce the highest objective function. On the other hand, when the Pc values of 0.55 and 0.95 are used, the algorithm cannot achieve the highest value of the objective function with the given computation period.

Similarly to the tests with Pc, different Pm values cause different levels of performance of GA-AS. With the Pm of 0.05 and 0.15, the algorithm successfully found the solution with the highest net total benefit. The algorithm, on the other hand, cannot find the optimal solution within the computational period given when the Pm of 0.25 and 0.35 are adopted. The results show that the performance of the GA-AS algorithm significantly relies on the values of both Pc and Pm.

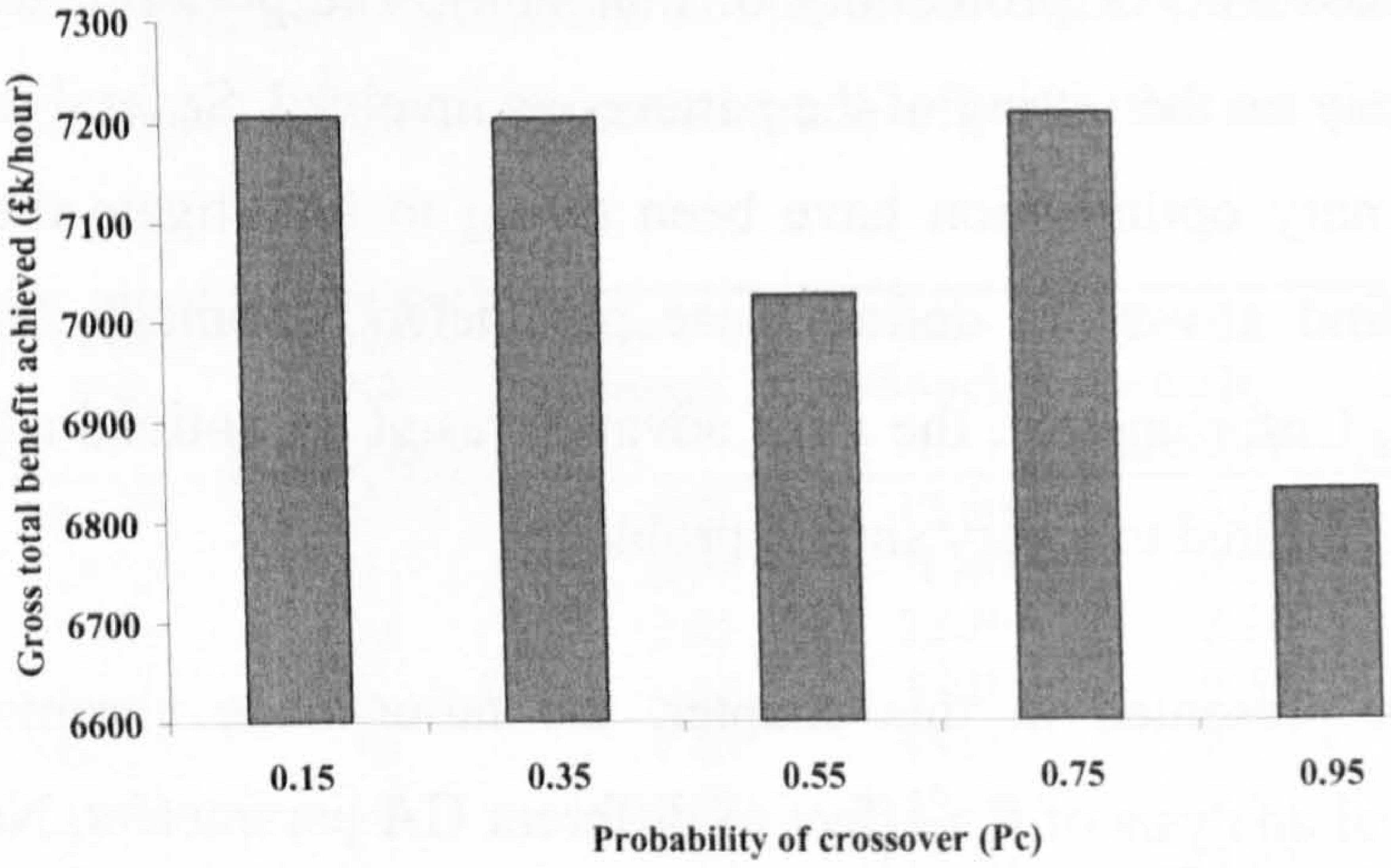


Figure 6-27 Comparison of the performance of GA-AS with different values of probability of crossover (Pc)

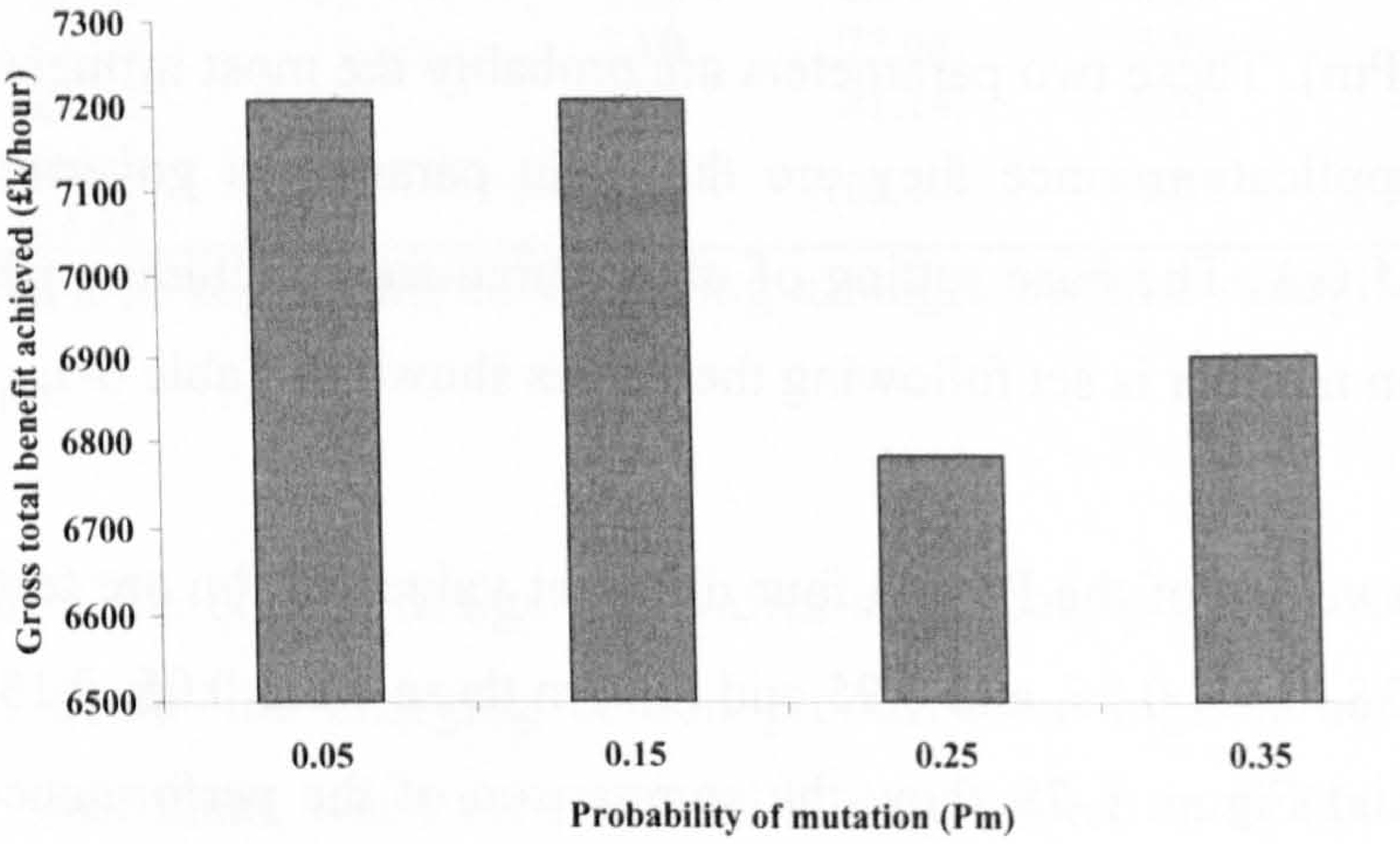


Figure 6-28 Comparison of the performance of GA-AS with different values of probability of mutation (Pm)

The other indicator which should be looked at when considering the performance of the GA algorithm is the speed of the convergence of the algorithm. Figure 6-29 and Figure 6-30 show the profile of the development of the best fitness values found in each generation of the algorithm for the different values of Pc and Pm respectively. These profiles provide additional information on how quickly the algorithm converges to the final solution. From Figure 6-29, the test with Pc = 0.75 converges to the final solution fastest (just after about 90 generations). The test with Pc = 0.15 comes second in which the algorithm converges to the final solution after about 120 generations. The test with Pc = 0.35 converges to the final solution after about 180 generations. As explained earlier, the tests with Pc = 0.55 and 0.95 do not converge to the best solution found by the other three tests.

In a similar way, the results in Figure 6-30 show the effect of the Pm on the speed of the convergence of the GA-AS. Based on the figure, the test with Pm = 0.05 converge to the solution fastest (after about 50 generations). The test with Pm = 0.15 (which is the base setting) comes second and the other two tests (with Pm = 0.25 and 0.35) do not converge to the final solution.

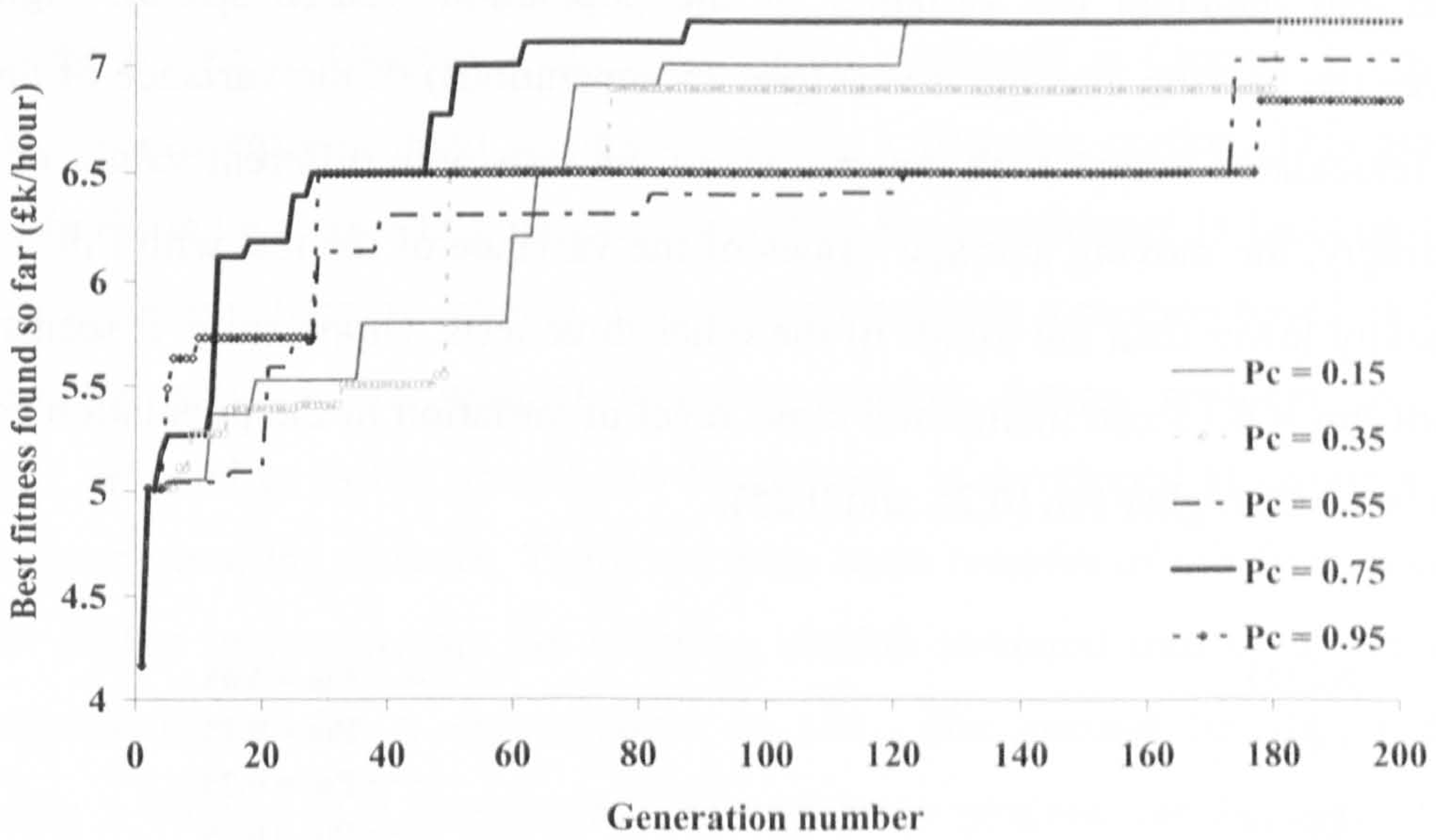


Figure 6-29 Profile of the improvement of the best solution from the GA-AS with the different values of Pc

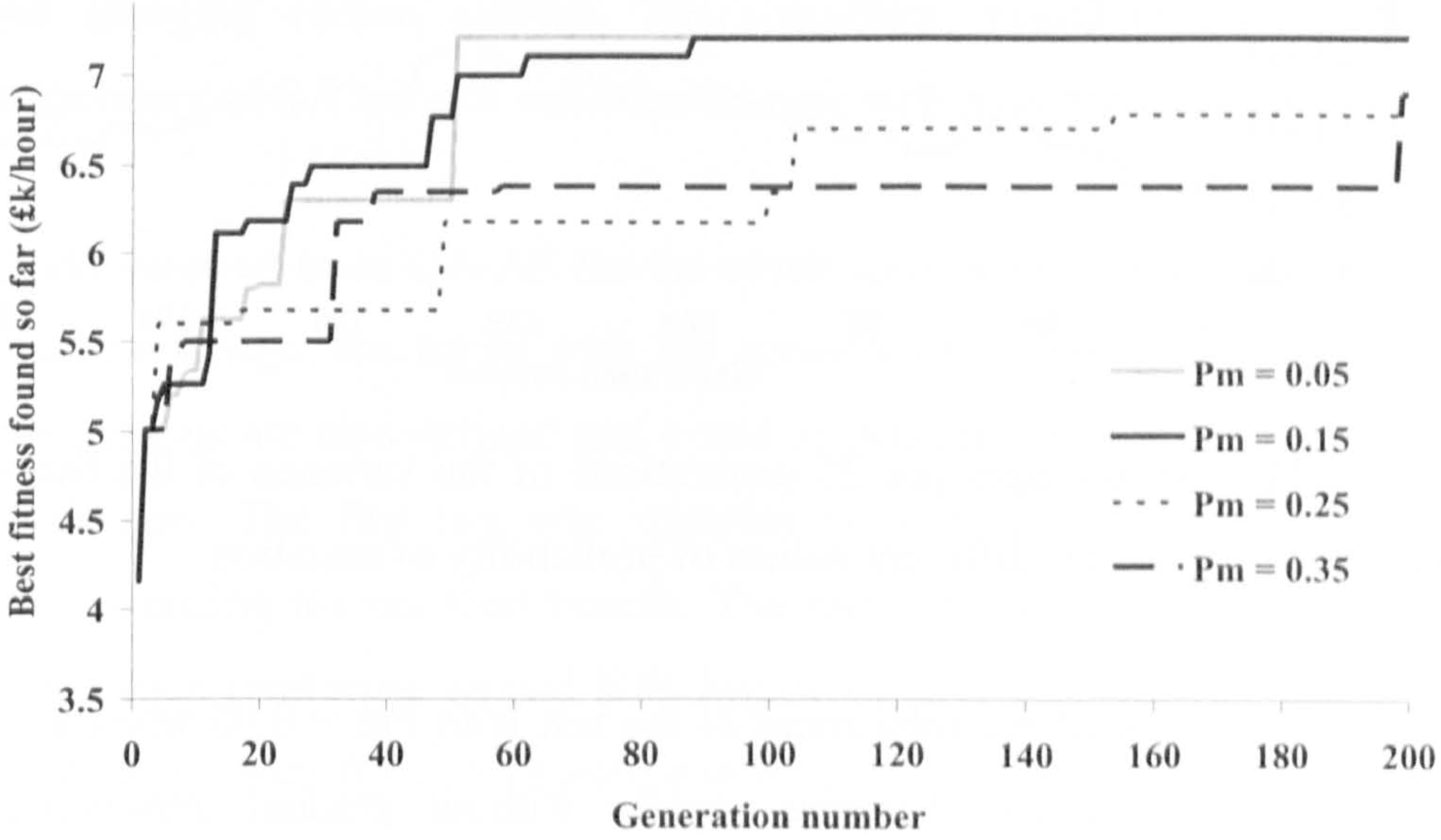


Figure 6-30 Profile of the improvement of the best solution from the GA-AS with the different values of Pm

Despite the quick convergence of the test with Pm = 0.05, a lack of variation of the population could be detected. From the profile of the fitness improvement of the test with Pm = 0.05, the best value of the net total benefit increases rapidly from the fitness just under £6k to the final solution (fitness value of around £7k) with only two improvement steps (the first improvement is at about generation 25 and the second is at about generation 45). A good parameter set up for GA should converge to a good solution and maintain the variation of the population (search space). Figure 6-31 compares the moving average trends (per 25 generations) of the variance of the fitness values (calculated within each generation) of the tests with different values of the Pm. Interestingly, the moving average values of the variance of the test with Pm = 0.05 are consistently lower than the trends of the other three tests. Graphically, it seems that the test with Pm = 0.15 can maintain a close level of variation in the population compared to the tests with higher Pm (0.25 and 0.25).

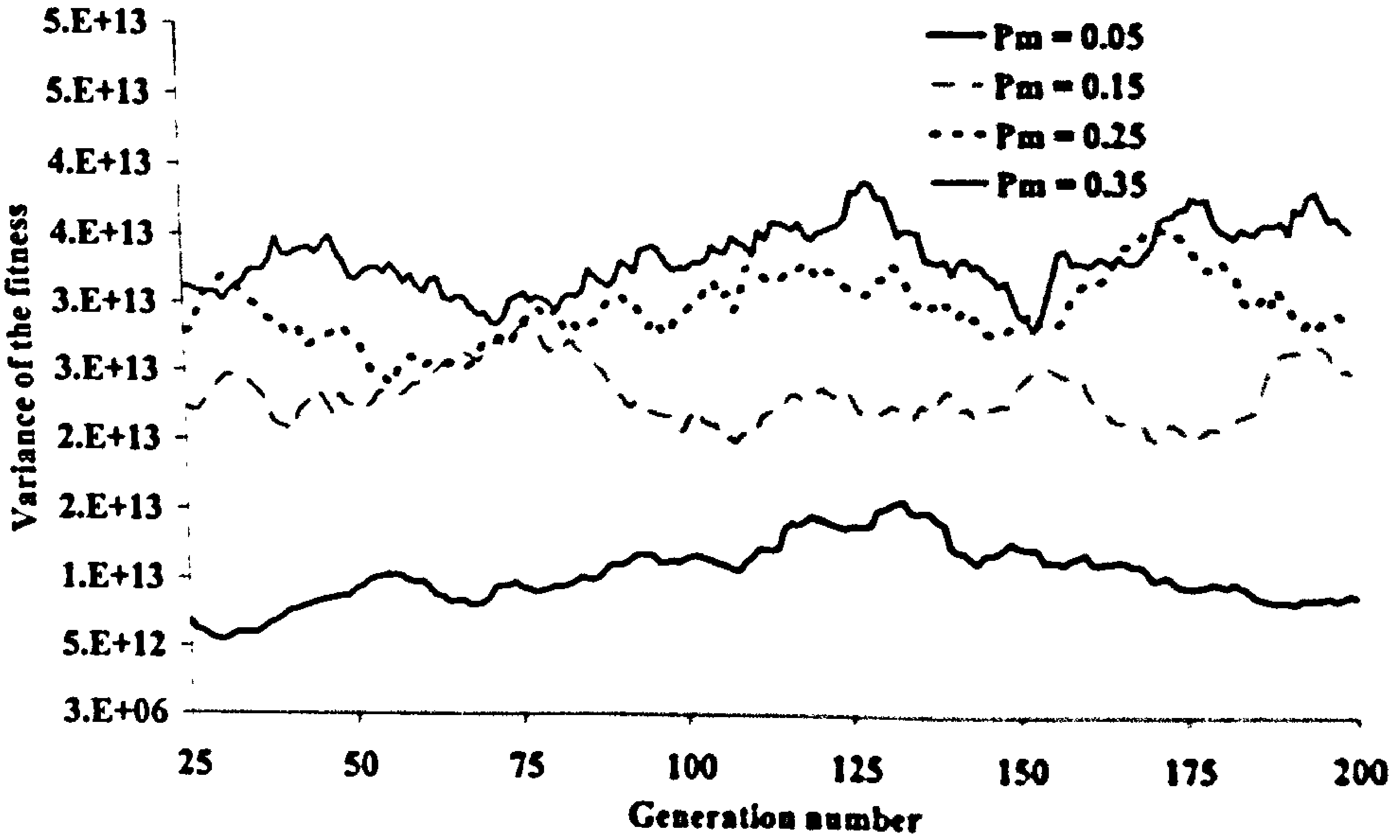


Figure 6-31 Moving averages per 25 generations of the variance of the fitness values within each generation for different values of probability of mutation

It is a sign of the lack of the robustness of the test with Pm = 0.05 when a significant improvement of the fitness happens rapidly without gradual improvement. This phenomenon coincides with the low level of the variation of the population (compared to the other three tests). The gradual improvement (which is the trend of the test with Pm = 0.15) can happen when the key improvement of the chromosome comes from the crossover operator of GA. This will happen when a variety of population is available within each generation. On the other hand, if there is a little variety of population, the ability of the crossover to devise an improved solution decreases. In this case, the main mechanism to find a better solution is left to the mutation process. This is rather a 'hit or miss' searching process which could not guarantee the robustness of the algorithm.

6.7 SUMMARY

Based on the review in Chapter 2 and the findings presented in Chapter 3, one of the most practical designs of a road pricing scheme is a charging cordon. This presents a challenge in devising a sounding method for aiding the practitioner in locating the best location of a cordon charging scheme. This chapter presents a method based on Genetic Algorithm for solving the optimal charging cordon design problem. A special framework, referred to as the branch-tree framework, is developed to define a cordon from the structure of the network. There are three main benefits of this framework. The first is its ability to encapsulate the charging cordon structure into string which is a typical representation for a chromosome in GA. The second benefit is that by representing a cordon as a branch-tree, the crossover process can be applied to the chromosome directly and the resulting chromosome is ensured to be a closed charging cordon. The last benefit is that it allows the mutation process to be applied so that a mutated chromosome is a new cordon. The framework is also extended to the case for a multi-layer charging cordon scheme. The properties related to the crossover and mutation operators of GA are still valid for the case with a multiple cordon.

The methods, referred to as GA-AS for the single cordon design and GA-ASII for the multiple cordon design, are tested with the network of Edinburgh city of UK. Three judgmental cordons are also defined and tested to compare their performances with the optimised designs. The first test was applying GA-AS to find an optimal charging cordon to maximising the net total benefit. The resulting design outperformed the best judgmental cordon producing around 80% higher benefit. GA-AS was also adopted to optimise the design with the new objective to maximise the net revenue. This is mainly to demonstrate the flexibility of the algorithm for dealing with different objective functions.

According to the survey results in Chapter 3, the toll level of the charging cordon scheme may be defined in advance for some political reasons. Thus, some tests were conducted by applying GA-AS to find an optimal cordon location for a given uniform toll level. Six different toll levels were tested (£0.50, £0.75, £1.00, £.125, £2.00, £3.00, and £4.00). The results show that with the toll level higher than £3, the optimised design cannot generate any positive benefit. By restricting the toll level which is not the optimal one (the optimal toll for OPC1 is found to be £1.50), the benefit of the optimised design decreases. This represents an interesting possible trade-off between the practicality of the toll level and the benefit of the scheme. GA-ASII was used to find an optimal location for double charging cordons. By allowing a higher number of charging cordons, the optimised double cordon scheme generate a substantial higher benefit compared to the benefit from OPC1.

Apart from the illustration of the performance of the methods, a test was conducted to compare different searching strategies within the GA-AS algorithm. The comparison was to evaluate the best approach to optimise the toll level and toll location simultaneously. The first strategy tested is the grid-search strategy where each cordon is tested with a different predefined uniform toll. The toll level producing the highest benefit is assigned as the optimal toll level for that cordon and the associated benefit is assigned as the fitness of the chromosome. In this approach, for a chromosome a number of objective function evaluations (equal to the number of predefined toll levels) must be conducted to determine the fitness of that chromosome.

The alternative strategy tested is the toll vector strategy where for each cordon (chromosome) a toll level is attached to the design in the form of binary string. Both the cordon location and toll level are evolved simultaneously. From the result, it is found that the toll vector strategy outperformed the grid-search strategy. The main reason is thought to be a wider searching space for the toll vector approach where all chromosomes with different toll levels are pooled for the crossover operation. With this approach, despite the same number of modelling runs, the crossover operator is applied to a higher number of chromosomes compared to the mechanism of the gird-search strategy.

The other performance test is related to the robustness of the algorithm based on different parameter settings of GA. Two GA parameters were tests including the probability of crossover (Pc) and the probability of mutation (Pm). Two indicators of the optimisation process were observed, the level of optimality of the solution and the speed of the convergence of the optimisation process. The results show that the performance of the GA-AS can be very sensitive to the different values of Pc and Pm. For the tests with Pc, three out of five tests (five values of Pc) achieved the highest objective function found so far. However, the speed of the convergence of each test varied. The Pc value of 0.75 came out best for both performance indicators.

For the tests with Pm, two out of four tests found the solution with the highest objective function (Pm = 0.05 and 0.15). The speed of the convergence of the test with Pm = 0.05 is slightly faster than the test with Pm = 0.15. One observation on the rapid improvement of the objective function of the test with Pm = 0.05 was made. This is one of unwanted characteristics for a robust GA. A closer investigation was made on the variation of the population. The result shows a low level of variation of the population for the test with Pm = 0.05. The rapid improvement of the objective function is thought to come mainly from the mutation operator ('hit or miss') whereas a good GA set up should rely more on the crossover operator. The crossover operator can work best with the certain level of the diversity of the population.

The results from various tests with GA-AS and GA-ASII aim mainly to demonstrate the performance and applicability of the method with a real network. The main specification for the problem in this chapter is to optimise a single objective with different conditions (e.g. predefined toll level or double cordon case). The test with the predefined uniform toll level shows an interesting trade-off between the constraint on the design and its main objective. Similarly, the tests with the different objective functions (i.e. with the objective functions of net total benefit and net revenues) reveal some possible conflict between different objectives of the scheme design.

As explained in Chapter 3, there are various design constraint/criteria on the actual scheme design (e.g. the acceptable level of possible equity impact, the required lowest level of revenue generated, or the maximum adverse impact of the scheme). Similarly, it is also found that a road pricing scheme is regularly aimed to serve several objectives.

In the first case, we need to extend the methods developed in this chapter (GA-AS and GA-ASII) to deal with a number of practical constraints. Similarly, to serve the second requirement the algorithm must be able to solve the design problem with multiple objective functions. Fortunately, the framework of the evolutionary optimisation algorithm of GA-AS allows the additional requirements to be integrated to the methods developed in this chapter. Chapter 7 and Chapter 8 explain the extension of the GA-AS method to deal with the constrained design and multi-objective design problems of a charging cordon scheme respectively.

CHAPTER 7 OPTIMAL CHARGING CORDON DESIGN WITH CONSTRAINTS

I saw the angel in the marble and carved until I set him free. Michaelangelo

7.1 INTRODUCTION

The survey with the local authorities reported in Chapter 3 identified several important requirements for a road pricing scheme design. It was found that the design of a charging cordon is regularly associated with a number of outcome constraints (e.g. minimum level of revenue generated, acceptable level of equity impact, reduction in congestion, etc.).

This chapter aims to extend the GA-AS method developed in the previous chapter to be able to deal with outcome constraints of the scheme. In particular, we consider four different outcome constraints in this Chapter: revenue, equity impact (as measured by the *Gini* coefficient), total travel time, and total travel distance. The constrained charging cordon design problem is referred to as CON-OPC (Constrained Optimisation Problem for charging Cordon design) and in this chapter a self-adaptive penalty-based algorithm is applied to the GA-AS to solve the CON-OPC problem. The rest of the chapter is structured as follows:

- Section 7.2 reviews the approach for dealing with the constrains in GA. The methods are categorised into three main categories based upon their strategies including the constraint relaxation method, interior search method, and hybrid method.
- Section 7.3 then explains the self-adaptive penalty based algorithm that will be implemented with the GA-AS for solving CON-OPC. The integrated method is named CON-GAAS.
- The CON-GAAS is then tested in Section 7.4 with the network of Edinburgh. The other two possible penalty based approaches (as reviewed in Section 7.2) including the static and dynamic penalty based approaches are also tested to compare the behaviour of the penalty function and GA process with the self-

adaptive method. Four constraints are considered: the minimum level of revenues, maximum level of equity impact measured by the *Gini* coefficient (see 4.4), total travel distance, and total travel time. Note that similar to the previous chapter the discussion of the results in this section is mainly to demonstrate the approach. The policy implications of the results and analysis of the impact will be discussed more fully in Chapter 9.

- Finally, Section 7.5 summarises the chapter.

7.2 GA APPROACHES FOR CONSTRAINED OPTIMISATION PROBLEMS

This section reviews possible methods to deal with constraints in GA. Of course, there are many ways to categorise all possible strategies for handling the constraints in GA (See Coello, 2002, for a very comprehensive review on the same topic). The methods are categorised into three groups: constraint relaxation method, restricting search inside the feasible region, and hybrid method. The category presented in this section is only one of many possible ways. The category of the approach is constructed by referring back to the approach for dealing with the constraints in the conventional optimisation theory.

In conventional optimisation theory, there are two main ways for dealing with the constraints in the optimisation problem. The first is to relax the constraint allowing the initial solution to be outside the feasible region and then gradually force the solution to the optimal and feasible region (this is referred to as the exterior penalty function method). The second strategy is to start off the search inside the feasible region and ensure that the search solutions are kept inside the feasible region (e.g. barrier function method, simplex, or projection method). Despite their differences, both approaches rely on the same concept that is to transform the constrained optimisation problem to an unconstrained one.

The main two strategies for handling the constraints adopted in GA or, in a more general context, Evolutionary Algorithms (EA) are similar to those adopted in the conventional optimisation methods. The first strategy, referred to as a constraint relaxation method, is to separate the task of optimising the main objective function from the task of ensuring the feasibility of the solution. Different methods in this category adopt different degrees of separation. The second strategy is to keep the solutions feasible during the search. Apart from these two strategies, which are mapped very well with the conventional optimisation methods, the fast development of GA and EA also creates various new ways for handling the constraints. These methods are referred to as the *'hybrid method'*. The next three sub-sections will go through each of these three constraint handling strategies in GA.

7.2.1 Constraint relaxation method

Different constraint relaxation approaches with GA have been proposed for tackling the constrained optimisation problem including exterior penalty based method (Homaifar *et al*, 1994; Joines and Houck, 1994; Hadji-Alouane and Bean, 1997), co-evolution (Paredis, 1994), behavioural memory (Schoenauer and Xanthakis, 1993), and multiobjective optimisation (Surry and Radcliffe, 1997). The first strategy proposed to handle constraint in GA is the penalty based method. This method allows the infeasible chromosomes to be included in the population and survive to the next generation. The method then gradually forces the searching mechanism to eliminate infeasible chromosomes and to converge to a set of feasible solutions. This type of strategy is similar to the exterior penalty based method in the conventional optimisation algorithm (Bazaraa *et al*, 1993).

The constrained optimisation problem is transformed to an unconstrained one by creating an auxiliary function. Let $\psi(x)$ be the objective function of the optimisation problem (say a maximisation problem) and let $g_i(x) > 0$ for $i = 1 \ldots m$ be the constraints of the problem. The auxiliary function representing the fitness of the chromosome can be constructed as follows:

$$\hat{\psi}(x) = \psi(x) - \sum_{i=1}^{m}\left(penalty \cdot \left|\min\left[g_i(x), 0\right]\right|\right) \qquad \text{Equation 7-1}$$

where $\hat{\psi}(x)$ is the auxiliary fitness function for the chromosome and *penalty* is the penalty applied to fitness of the chromosome if the constraints are violated.

The penalty function plays a crucial role in reconciling the attempt to maximise the original objective function and the effort in finding feasible solutions. The adjustment of the penalty level is crucial to the success of the algorithm (Davis, 1987; Richardson *et al*, 1989). Assigning too high a penalty causes the search to lose useful information from the infeasible chromosomes whereas using too low a penalty makes the search waste too much time in infeasible regions (Crossley and Williams, 1997; Coello, 2000c; Deb, 2000).

Various schemes for defining the penalty term have been proposed.

- *Static penalty*: Homaifar *et al* (1994) proposed the static penalty function that does not change with the generation number. The penalty can be defined so that a higher penalty is applied to the chromosome with a higher level of constraint violation. In most cases, the user defines the penalty term in Equation 7-1 as a set of constants ($K_{i,r}$) for each constraint i and each level of violation r (also defined by the user). Other possible forms for the static penalty function are also proposed (see Hoffmeister and Sprave, 1996; Kuri Morales and Quezada, 1998).
- *Dynamic penalty*: The dynamic penalty function is one of the alternatives to the static version. In contrast to the static penalty function, the dynamic penalty function is varied from generation to generation, i.e. the generation number is one of the arguments in the penalty function. Joines and Houck (1994) proposed a simple approach to defining the dynamic penalty function in which the penalty term increases with the generation of GA. A more sophisticated form of the dynamic penalty function was also proposed by Michalewicz and Attia (1994) based on the idea of simulated annealing.
- *Self-adaptive penalty*: The other class of dynamic penalty function is the self-adaptive penalty function (Richardson *et al*, 1989; Hadji-Alouane and Bean, 1997). This type of penalty function incorporates the feedback from the search process into the determination of the appropriate penalty value. Examples of the types of feedback adopted in the dynamic penalty function include the feasibility of the best individual in the previous generations or in the last generation (Hadji-Alouane and Bean, 1997; Rasheed, 1998), average fitness of the population in the previous generation (Richardson *et al*, 1989), and the best-known fitness overall before being penalised (Smith and Tate, 1993; Coit *et al*, 1996).

The self adaptive dynamic penalty function is reported as a superior approach for constrained optimisation problem out of these three methods (Crossley and Williams, 1997). However, most of the tests conducted are inconclusive and the performance of the self dynamic penalty function also depends significantly on problem specific characteristic.

The penalty based methods (static, dynamic, and self-adaptive) have been influenced largely by the framework of the conventional optimisation theory. GA or EA in fact operate on a more flexible optimisation paradigm and are both population-based search algorithms. These entities allow various innovative constraint relaxation approaches to be implemented with GA and EA. The strategy adopted in the penalty based method includes all constraints into the evaluation process simultaneously. All constraints are evaluated and influence the searching mechanism in the same time. The other possible strategy proposed is to deal with each constraint sequentially or separately. Two possible methods steming from this idea are *behavioural memory* (Schoenauer and Xanthakis, 1993) and *co-evolution* (Paredis, 1994).

The behavioural memory method splits the task of optimising the main objective function from the task of finding feasible solutions. This method is similar to the approach called *lexicographic ordering* commonly used in multiobjective optimisation (Coello, 1999b). During the evolution in iteration t with constraint t, a chromosome which does not satisfy the active constraint in the iteration $(t-1,\ t-2,\dots,1)$ will be eliminated from the population.

Paredis (1994) proposed a technique based on a co-evolutionary model inspired by the predator-prey model. The constraints are treated as another group of the population. The chromosomes representing a solution and a constraint are randomly drawn from two population groups (in tournament selection fashion). If the chromosome violates the encountered constraint, that chromosome receives a penalty. Otherwise, the chromosome receives a reward. The encountered constraint gets a penalty when the chromosome satisfies that constraint and gets a reward otherwise. The chromosome and constraint with a higher fitness has a higher probability to be selected for an encounter.

The strategies described above mainly treat the constraint as the conventional optimisation theory does. Recently, some research in the area of EA and GA has proposed a new way of treating the constraint as one of the objective functions (Camponogara and Talukdar, 1997; Surry and Radcliffe, 1997; Coello, 2000a; Coello, 2000b; Ray, 2002). The main idea is to redefine the constrained single objective optimisation problem as a multiobjective optimisation problem consisting of $m+1$ objectives (where m is the number of constraints). Then a number of possible EA and GA methods for dealing with the multiobjective optimisation problem can be adopted (See Deb, 2001).

7.2.2 Restricting search inside the feasible region

Most of the methods for dealing with the constraints described in the previous section allocate sufficient level of computation effort and memory searching through the infeasible region with the hope that the information from the infeasible chromosomes will lead the search to the optimal feasible chromosome. The methods from the constraint relaxation family may lose this efficiency if the problem is highly constrained and the task of finding a feasible solution is as important as finding the optimal solution. In particular, the problem with constraints on the actual structure of the solution (e.g. constraint on the feasible combination of variables) poses a serious challenge to the application of GA and EA to the constrained optimisation problem

An alternative approach to the constraint relaxation method is to restrict the search within the feasible region. This approach remedies the drawback of the constraint relaxation method when dealing with a highly constrained problem. Various approaches have been proposed for carrying out the operation, including using a special representation and/or tailor-made genetic operators -crossover and mutation- (Davis, 1991), mapping between the encoded chromosomes and feasible search space (Bean, 1994), approximating the boundary of the feasible region (Schoenauer and Michalewicz, 1997), and repair algorithm (Liepins and Vose, 1990).

The most natural way to restrict the search within the feasible region is to design a special representation of the chromosome and/or the genetic operators (crossover and mutation) to ensure the feasibility of all chromosomes (Davis, 1991). Gen and Cheng

(2000) described this approach as the development of the adapted GA to suit a more complex real world problem. Some examples of the application of the representation approach includes the design of optimal charging cordons described in the previous chapter and as proposed in Yang *et al*, (2002), factory layout problem (Suresh *et al*, 1995; Gomez *et al*, 2003), supply chain network design (Jang *et al*, 2002; Zhou *et al*, 2002), multistage process planning problems (Gen *et al*, 2001), air traffic control planning (Hansen, 2004), and transport network topology design problems (Drezner and Wesolowsky, 2003). In many cases, the genetic operators (crossover and mutation) must also be redefined to suit the adapted chromosome structure and to sustain the feasibility of off-spring after applying the genetic operators to the parent chromosomes.

Interestingly, most of the problems adopting the special representation approach as the strategy are network related problems. This is intuitive because the topology requirement (from the structure of the network) of the solution inherited in the constraint may be too difficult for the constraint relaxation method to handle. With the special representation and genetic operators, all chromosomes are ensured to be feasible whereas the algorithm with the constraint relaxation method may spend most of its computational time to just find a feasible solution.

The other approach to restrict the search inside the feasible region is to define a mapping between the normal chromosome representation and the feasible solution space. This can be done through a special decoder (Koziel and Michalewicz, 1998). A somewhat different paradigm for keeping the search inside the feasible region is to repair the infeasible chromosome to be a feasible one, referred to as the repair algorithm (Liepins and Vose, 1990). This technique is adopted widely in the area of combinatorial optimisation where it is relatively easy to repair the infeasible solution.

7.2.3 Hybrid methods

The last category of constraint handling methods is the hybrid method. The methods in this category are mainly coupled with other optimisation techniques including conventional and meta-heuristic methods. These include the lagrangian penalty based method, random evolution, fuzzy logic with EA, the immune system based method, and cultural algorithm. The detail of these algorithms will not be discussed here. However, it should be noted that most of these methods can also be put into one of the constraint handling categories explained earlier. For example, Coello (1999a) Suggested that the random evolution method can also be considered as one kind of the repair algorithm. Similarly, the lagrangian penalty and fuzzy logic based method can also be put into the category of penalty based methods.

7.3 CONSTRAINED CHARGING CORDON DESIGN: SELF-ADAPTIVE PENALTY METHOD

An advantage of GA in dealing with an arbitrary formulation has already been exploited in Chapter 6 in dealing with the topological constraint for the set of tolled points. The branch-tree approach proposed in Chapter 6 provided a way to deal with the constraint using a special representation of the chromosome. In this chapter, additional constraints on the cordon design will be included into the optimisation process. The constraints mainly involve different performance indicators of the scheme design (net revenue, equity impact, total travel time, or total travel distance).

These constraints are considered as *'hard constraints'* where the final solution must agree with the set constraint values. In this case, the approach based on multiobjective optimisation may not be appropriate since it cannot guarantee the feasibility of the solution at the end of the process. Most of the performance indicators involved in the constraints have to be calculated from the modelling output of the design test. As explained in Chapter 4, the structure of MPEC complicates the relationship between these output indicators and the design variables (i.e. cordon location and toll level). With this complex relationship, it is difficult or almost infeasible to develop a method to govern the search to be inside the feasible space either by special representation or using the repairing algorithm.

With this limitation, the mechanism adopted in this chapter to handle the constraints is the penalty based approach. This is a method in the class of constraint relaxation strategy as mentioned in Section 7.2. The method is able to handle the constraint related to the outcomes of the cordon design; it is also simple and can be easily integrated with the GA-AS method described in Chapter 6.

In this chapter, the method of "*dynamic self adaptive penalty*" proposed by Richardson, *et al* (1989) is adopted. The modified objective function taking account of the constraints $g_l \geq 0$ is as follows:

$$\hat{\psi}(x) = \psi(\mathrm{x}) - \left(\mu \cdot \left(\frac{gen}{Gen}\right)^{\rho} \cdot \bar{\psi}_{gen-1} \cdot \sum_{l} \left|\min(g_l, 0)\right|\right) \qquad \text{Equation 7-2}$$

where $\psi(\mathrm{x})$ is the original objective function, *Gen* is the total number of generations tested, *gen* is the current generation number, $\bar{\psi}_{gen-1}$ is the average fitness (without penalised) in the previous generation (*gen* -1), ρ and μ are parameters, and $\min(g_l, 0)$ returns "*the degree of constraint violation*" for constraint *k* (the formulation is for the constraint in the form of $g_l \geq 0$, but it can be easily changed to deal with $g_l \leq 0$ without lost of generality).

With this modified objective function, GA-AS can then deal with the design of the optimal charging cordon with its outcome constraints. Note that it is also relatively straightforward to implement other kinds of penalty function method as explained in Section 7.2 (i.e. static and dynamic penalty function). In Section 7.4, the behaviour of different penalty methods will also be compared using the Edinburgh network.

7.4 TESTS WITH THE EDINBURGH NETWORK

This section presents some tests with the GA based method for solving constrained charging cordon design. Two sets of tests are presented below. The first test, Section 7.4.1, is the comparison of the behaviour of different penalty based methods (i.e. static, dynamic, and self-adaptive). Although the main method adopted in this chapter is the self-adaptive penalty method, as explained in Section 7.3, it is of methodological interest to investigate the extent to which the formulation of the penalty function influences the GA searching process. The second test, Section 7.4.2, concentrates more on the application of the method to the cordon design problem. The tests are conducted with the network of Edinburgh as used in Chapter 6. Several constrained designs are solved with either one or two constraints. Note that although the tests in Section 7.4.2

only involve five different constraints (i.e. net revenue, equity, total travel time, and total travel distance), other constraints can also be implemented with the method proposed in this chapter since the GA based approach allows a very flexible functional form of the constraint.

7.4.1 Comparison between static, dynamic, and self-adaptive penalty methods

Implementing the tests

The test problem in this section is to find the location of a single charging cordon with optimal uniform toll for the Edinburgh network (See Figure 6-15 in Chapter 6). An additional constraint on the minimum level of net revenue generated is introduced. The net revenue must be greater than or equal to £50 k, i.e. net revenue $\geq$ £50k. Next the actual implementation of the static penalty and dynamic penalty are described.

The static penalty method as suggested in Michalewicz (1992) is adopted in this test. The formulation of the penalty function is as follows:

$$\hat{\psi}(x)=\psi(x)-\sum_{i=1}^{m}\left(R_{k,i}\cdot\left|\min\left[g_i(x),0\right]\right|\right) \qquad \text{Equation 7-3}$$

where $R_{k,i}$ are the penalty coefficients used, $k = 1,\ldots,l$, where l is the number of levels of violation defined by the user. In this test, three levels of violation are defined and the penalty coefficients associated with different violation levels are given in Table 7-1:

Level of violation	Penalty coefficient ($R_{k,i}$)
net revenue < £19,000	0.8
£19,000 $\leq$net revenue < £38,000	0.1
£38,000 $\leq$net revenue < £50,000	0.05

Table 7-1 Levels of violations and penalty coefficients for static penalty function

The dynamic penalty tested follows the formula suggested by Joines and Houck(1994):

$$\hat{\psi}(x)=\psi(x)-(C\times gen)^{\alpha}\sum_{i=1}^{m}\left(\left|\min\left[g_i(x),0\right]\right|^{\beta}\right) \qquad \text{Equation 7-4}$$

where C, α, and β are constant parameters set by the user (in this test C = 0.005, $\alpha = 1$, and $\beta = 1$), and *gen* is the generation number. With the set parameters, the implemented dynamic penalty coefficient is a linear function with the generation number.

The self-adaptive penalty function is implemented as explained in Section 7.3:

$$\hat{\psi}(x)=\psi(x)-\left(\mu\cdot\left(\frac{gen}{Gen}\right)^{\rho}\cdot\bar{\psi}_{gen-1}\cdot\sum_{i}\left|\min(g_i,0)\right|\right),$$

where ρ and μ are set to be 0.95 and 0.0000005 respectively.

Test results

Figure 7-1 to Figure 7-3 show the average net benefits and average net revenues with in each generation for the runs with static, dynamic, and self adaptive penalties respectively. All three methods converge to the same solution with the net benefit of around £7k per peak hour and net revenues of around £57k per peak hour (satisfying the constraint). The focus of the tests is to observe the behaviour of different penalty methods.

The trends of the average net benefits and net revenues in Figure 7-1 for the static penalty are relatively flat compared to the trends in Figure 7-2 and Figure 7-3 (for the dynamic and self-adaptive cases). The reason is that a constant static penalty is applied consistently throughout the evolution process for the first case. Therefore, the chromosomes violating the revenue constraint have low probabilities to survive from the early generations. On the other hand, the self-adaptive and dynamic penalty applied a gradually increased penalty to the evolution process. Hence, in early generations, the chromosomes with some constraint violation still have a good chance to survive (this explains the lower level of the average revenue in early generations in the cases of self-adaptive and dynamic penalty compared to static penalty). Then, as the evolution process proceeds, a higher level of penalty is gradually applied resulting in the form of increasing trend of the average net revenue as shown in Figure 7-2 and Figure 7-3.

The trends of the average net benefit and net revenue for the cases of dynamic and self adaptive penalty are very similar, i.e. the average net revenue gradually increases and the average net benefit gradually deceases. This is, as explained earlier, the intention of both dynamic and self adaptive penalty methods. However, from the actual formulation there should be a slight difference between these two methods. The trends of the average net revenue in Figure 7-2 seems to increases linearly with the generation number until around 50 generations (to about £45k). On the other hand, the trend of the average net revenue for the self-adaptive penalty increased to the level around £45k well before the 50th generation then settled down. These behaviours are consistent with the formulation of the penalty functions.

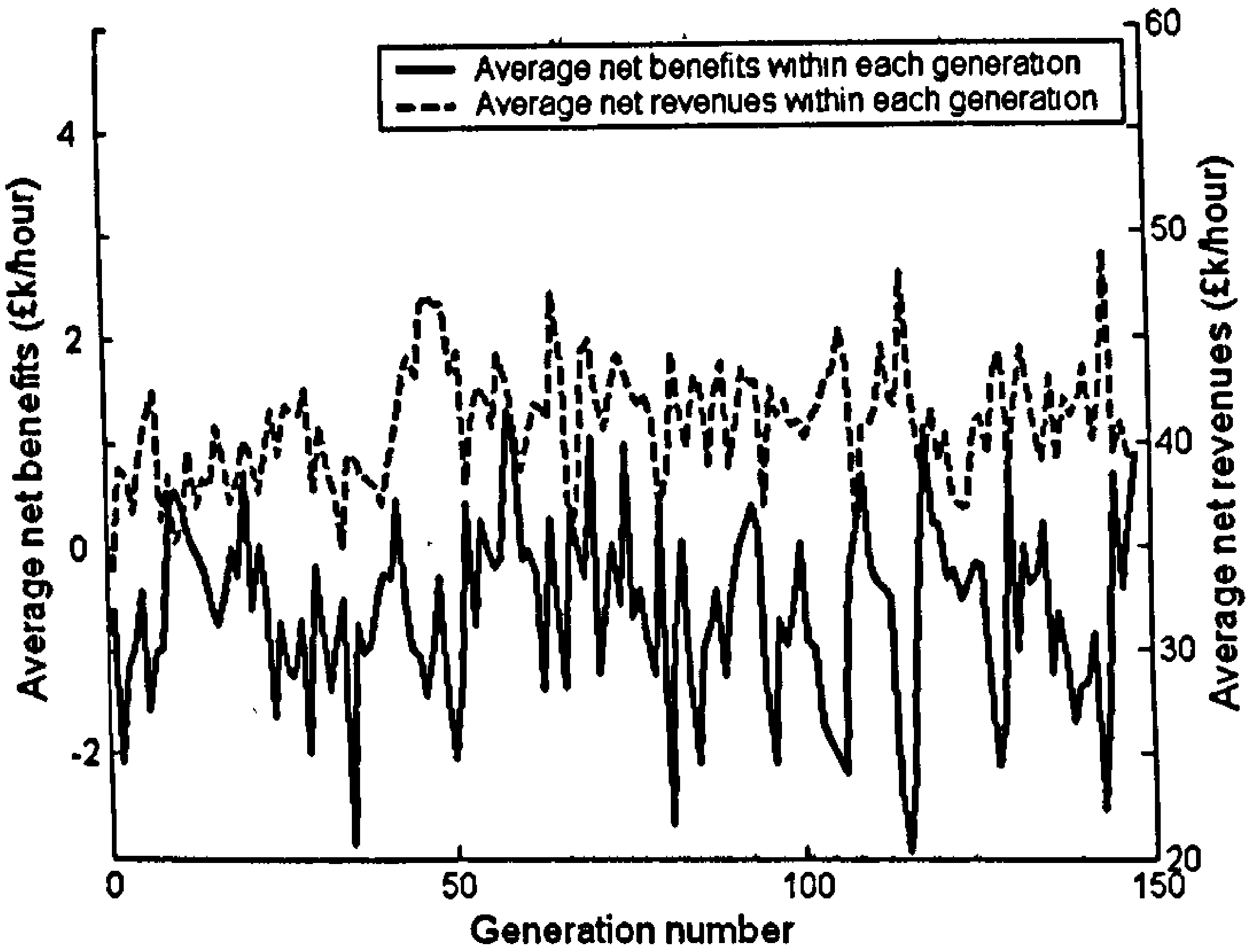


Figure 7-1 Average net benefits and net revenues within each generation (static penalty method)

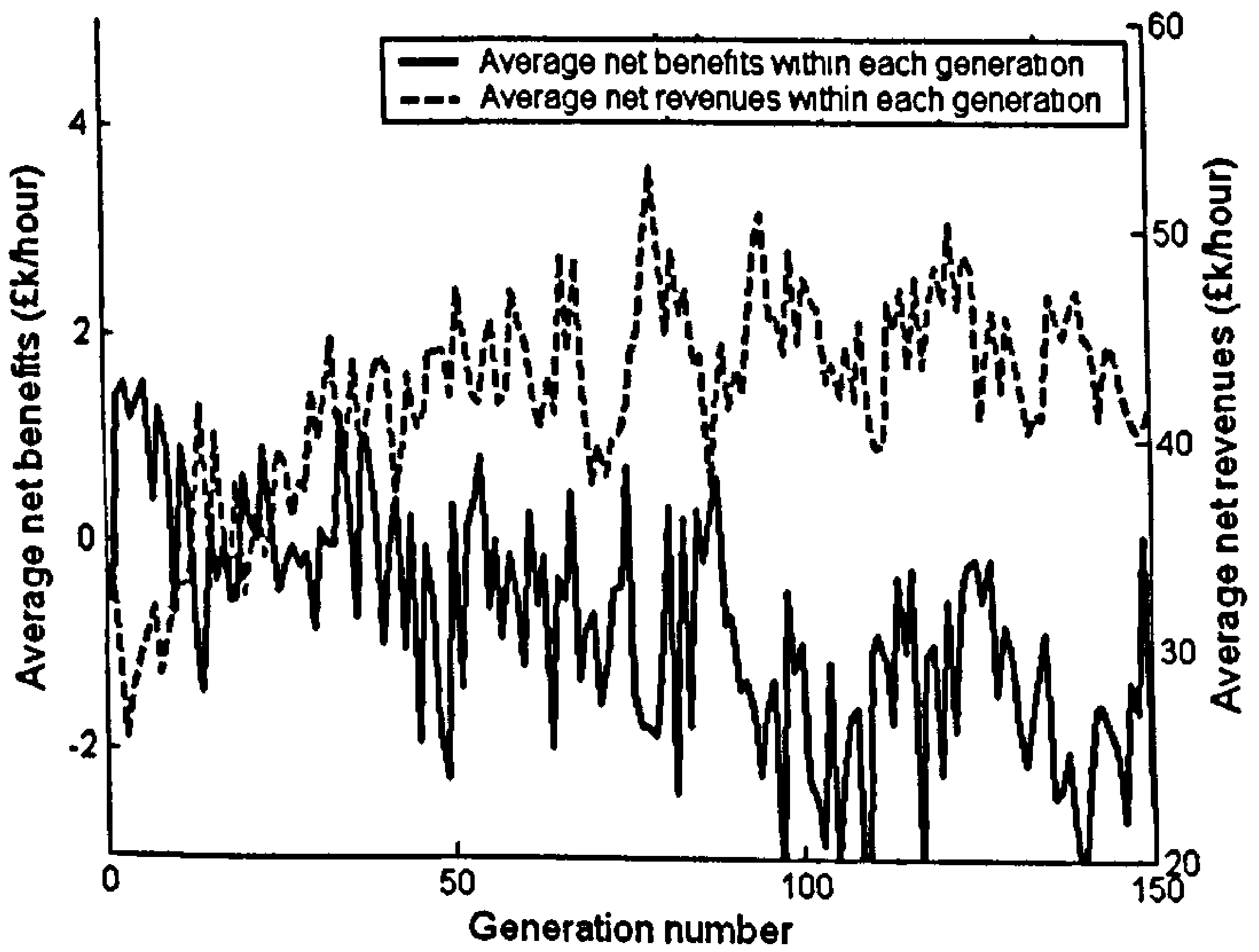


Figure 7-2 Average net benefits and net revenues within each generation (dynamic penalty method)

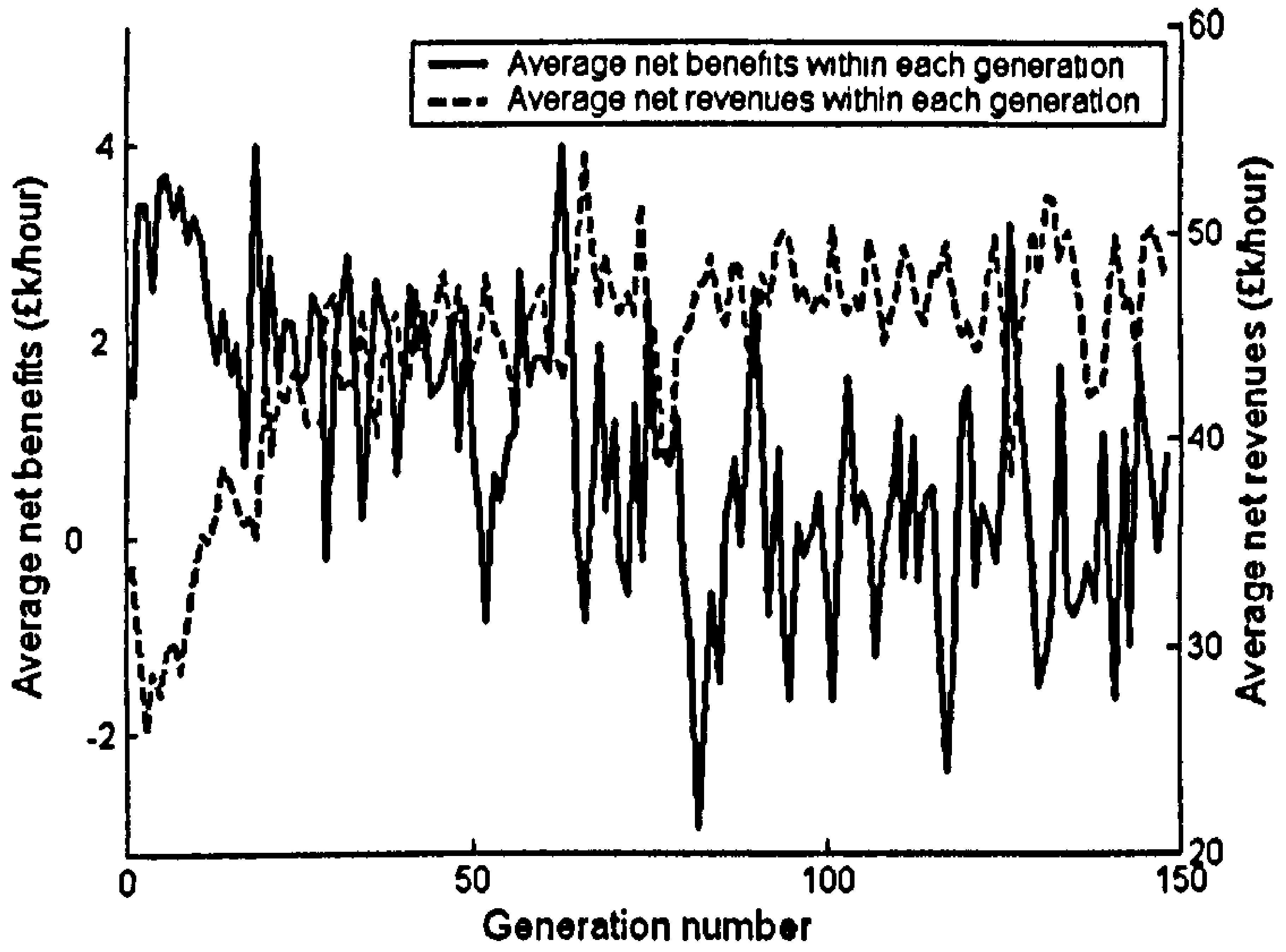


Figure 7-3 Average net benefits and net revenues (self-adaptive penalty)

For the dynamic penalty case, the penalty increases linearly with the generation number (regardless of the change in the average net benefit) pushing the average net revenue to around the level of the constraint set. On the contrary, the self-adaptive penalty also changes following the change in the average net benefit. In early generations, the low net average net benefit signals to the GA through the self-adaptive penalty to spend more effort on optimising the main objective function. This is reflected in the sharp increase in the average net benefit in Figure 7-3. The high level of the average net benefit then feeds back to the self-adaptive penalty in which the algorithm starts to highly penalise the infeasible solutions resulting in an increase in the average net revenues. This result, thus, explains the main difference between the dynamic and self-adaptive penalty methods.

7.4.2 Constrained charging cordon design

This section presents the results of the optimal cordon design with different constraints. Four different outcome constraints are included in the tests, including net revenue, the equity (distributional impact, see Section 4.4), total travel time, and total travel distance. Table 7-2 contains the performance indicators for the do-nothing scenarios and the tests with judgmental cordons which are used for setting up the tests. Five different tests are set up to illustrate the application of CON-GAAS:

(i) Maximise net benefit with revenue constraint (Net revenue ≥£50k)

(ii) Maximise net benefit with equity constraint (Gini ≤0.30)

(iii) Maximise net benefit with total travel time constraint (TTtime ≤49500 PCU-hr/hr)

(iv) Maximise net benefit with total travel distance constraint (TTdis ≤1690000 PCU-Km/hr)

(v) Maximise net benefit with equity and revenue constraints (Gini ≤0.30; Net revenue ≥£45k)

Cordon	**Optimal toll (£)**	**Net benefit (£k/hour)**	**Revenues (£k/hour)**	***Gini***	**TT travel time (PCU-Hr/hr)**	**TT travel distance (PCU-Km/hr)**	**Demand level (thousand PCU/hour)**
Do-nothing	-	-	-	0.12	58148	1780053	109.7
Inner 1	£0.50	2.10	11.71	0.26	56426	1767499	107.8
Inner 2	£0.75	3.99	17.70	0.31	55523	1767524	107.2
Outer	£0.75	1.96	20.20	0.20	55505	1734647	106.4

Table 7-2 Performance indicators for the do-nothing scenario and the tests with judgmental cordons

For the first test, the net revenue generated for the solution must be greater than or equal to £50k per peak hour (net revenue ≥£50k). This represents the fact that a road pricing scheme, to some extent, is used as the funding mechanism for improving other transport elements. The net revenue is that remaining after the scheme has financed its own operating and sunk costs.

As explained in Chapter 3, one of the most serious concerns with the charging cordon design is the distributional impact of the scheme. The second test is set up to represent this design scenario. Measuring the equity impact is a very complex task. In this test, the *Gini* coefficient explained in Section 4.4 is adopted as the possible measure of the equity impact. The constraint on the maximum acceptable level of equity impact (level of *Gini* coefficient) is introduced, *Gini* ≤0.30.

Similarly, another possible requirement imposed on the design of a charging cordon scheme discussed in Chapter 3 are reduction in congestion. Different measurements (indicators) for a reduction in congestion can be adopted with CON-GAAS, e.g. change in speed, or change in link delay. Also, the measurement of the reduction in congestion can also be focused on a particular part of the network, e.g. the central area. In this section, the measure of total travel time for the whole network is used as the proxy of the improvement of network condition in terms of the congestion reduction. The third test is set up to replicate this scenario where the required minimum level of travel time reduction is imposed onto the cordon design (*TTtime* ≤49500 PCU-hr/hr which is about 15% decrease in total travel time compared to the do-nothing case).

The last measure adopted for testing the CON-GAAS is the constraint on the total travel distance for the whole network. This measure can be used to as a proxy for different indicators that are directly related to the travel distance, e.g. some emission or accident costs. The fourth test is set up to illustrate the possibility of introducing the total travel distance as the constraint (*TTdis* ≤1690000 PCU-Km/hr). Finally, the fifth test is set up to show that the CON-GAAS can deal with a number of constraints simultaneously by including the constraints on both minimum level of net revenue and maximum level of equity impact.

The CON-GA is applied to all tests and Table 7-3 below shows the results for all the tests with the constrained designs. The final solutions found for the first and second problems are shown in Figure 7-4 and Figure 7-5 (named CON-REV and CON-GINI respectively). The summary of the key performance of the CON-REV and CON-GINI cordons is presented in Table 7-3. The optimal uniform tolls are £2 and £0.75 for CON-REV and CON-GINI. The net benefit from CON-REV is higher than that from CON-GINI; note that the net benefits from both designs are less than the net benefit generated by OPC1 found in Chapter 6. As expected, the revenue generated from CON-REV is around £57k per peak hour which satisfies the constraint set earlier. Similarly, the *Gini* coefficient of CON-GINI is 0.28 which is just below the acceptable criteria of 0.3.

Cordon	**Optimal toll (£)**	**No of toll points**	**Net benefit (£k/hour)**	**Net revenues (£k/hour)**	***Gini***	**TTtime (PCU-Hr/hr)**	**TTdis (PCU-Km/hr)**	**Demand (PCU/hour) '000**
OPC1	£1.50	13	7.21	43.70	0.41	52325	1720864	103.6
CON-REV	£2.00	13	6.99	56.44	0.40	51018	1703384	102.2
CON-GINI	£0.75	14	5.79	27.16	0.28	54328	1740770	105.4
CON-TTTIME	£3.00	13	3.75	75.74	0.39	48910	1673434	99.99
CON-TTDIS	£2.00	17	4.50	56.21	0.40	51229	1687498	101.64
CON-REV-GINI	£1.50	17	4.38	48.55	0.29	52139	1717133	102.5

Table 7-3 Overall results for constrained cordon designs

Figure 7-6 shows the Lorentz's curves (see Section 4.4, page 91 for the definition of this curve) for the benefit distributions of different charging cordons including OPC1, CON-REV, and CON-GINI. Note that the *Gini* coefficient is twice the area between the Lorentz's curve and the equality curve. Figure 7-7, Figure 7-8, and Figure 7-9 show the scatter plots of the user benefits for all O-D pairs for the OPC1, CON-REV, and CON-GINI cordons respectively. Interestingly, the effect of the equity constraint introduced in CON-GINI was the reduction of the user benefits for those O-D pairs enjoying a high level of benefit before in the OPC1 case. Simultaneously, the CON-GINI cordon also decreased the impact on those O-D pairs with negative user benefits.

The results imply that in order to include the design criteria on revenue and equity impact, the cordon design may need to trade-off some of the net benefit. From this particular set of tests, the constraint on revenue does not seem to significantly reduce the benefit of the scheme whereas the constraint of distributional impact (equity impact) does reduce the level of the net benefit substantially.

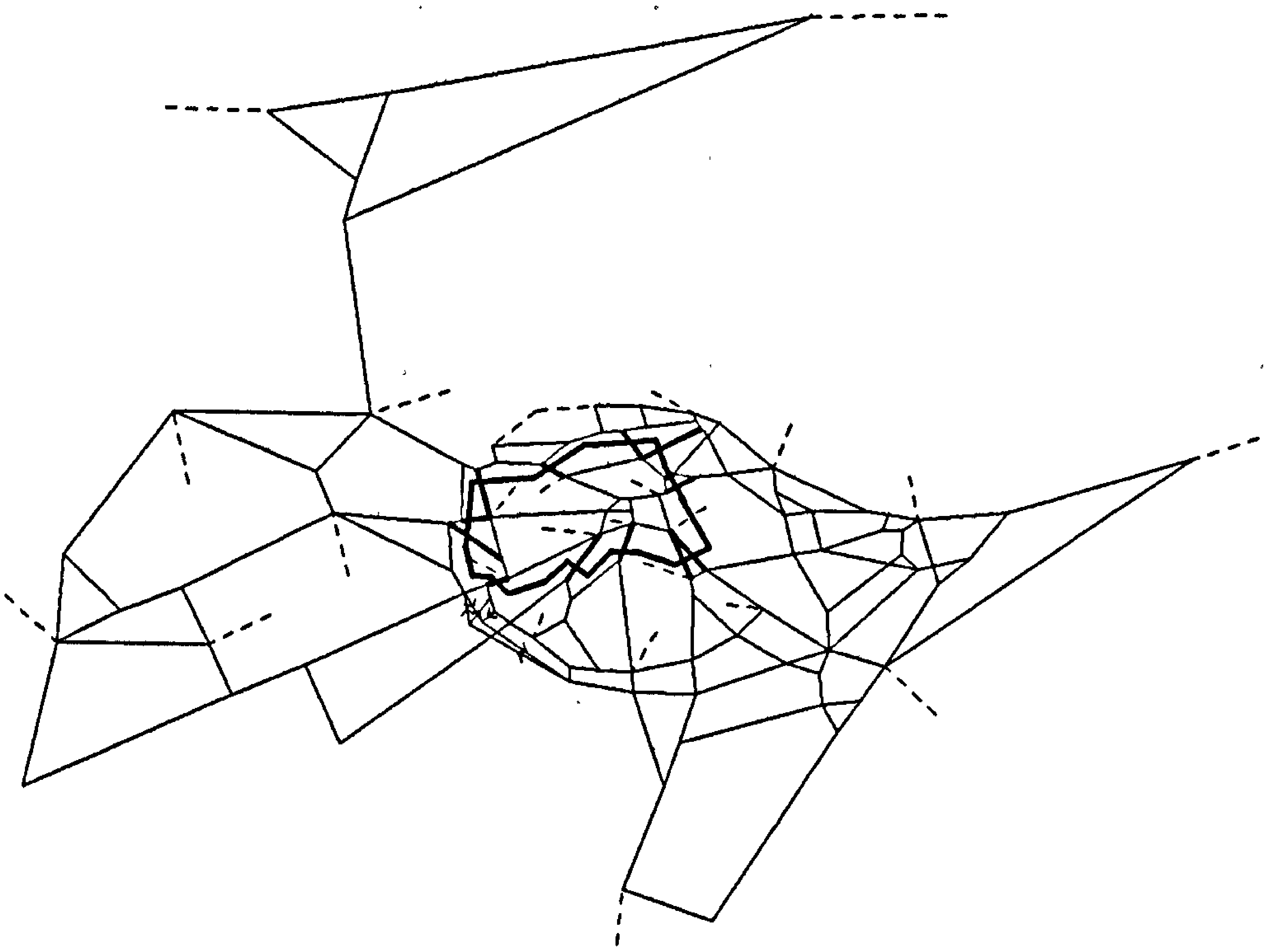


Figure 7-4 Optimal cordon location with constraint on net revenues (CON-REV) with toll level of £2.00

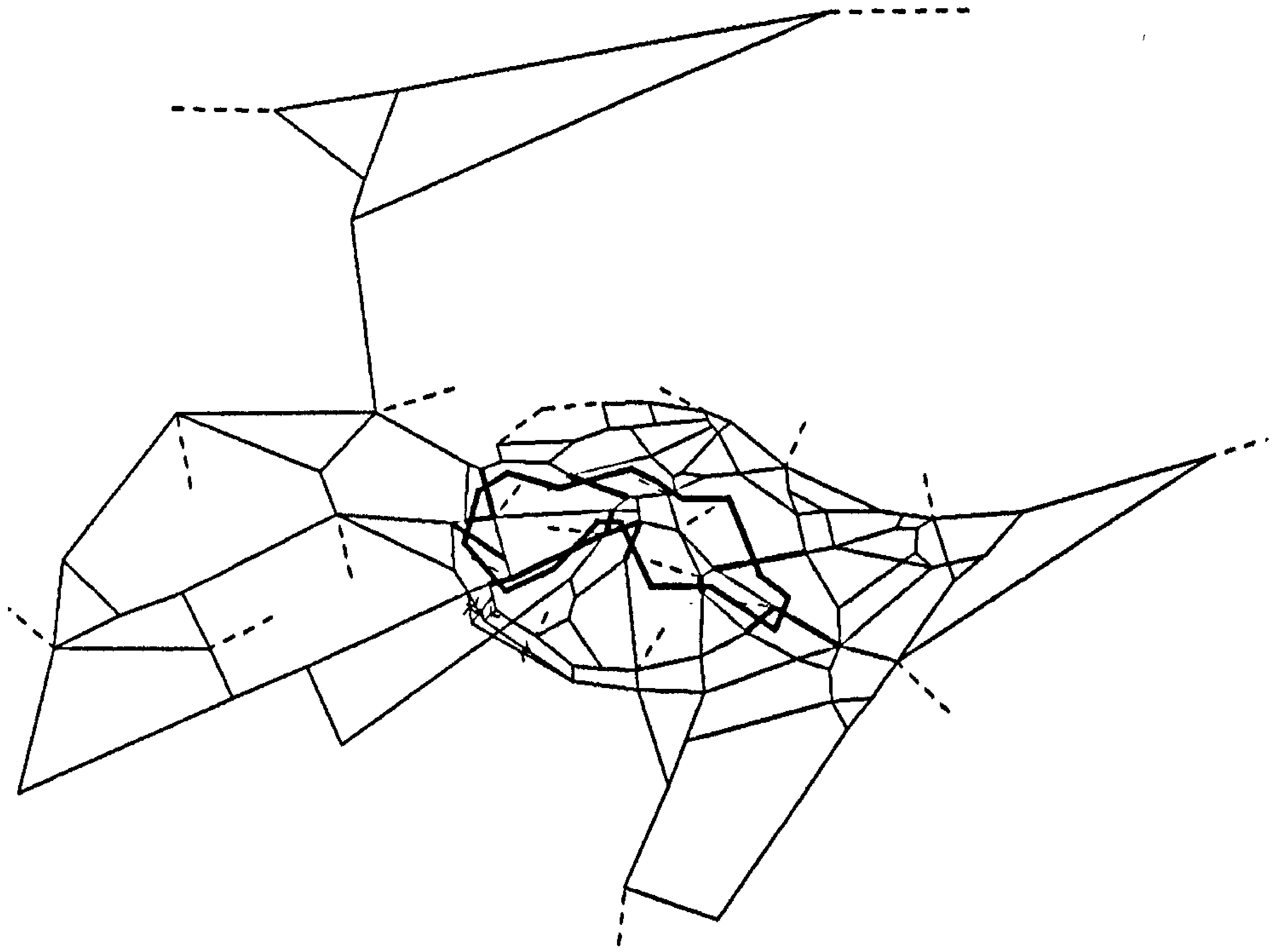


Figure 7-5 Optimal cordon location with constraint on equity impact (CON-GINI) with toll level of £0.75

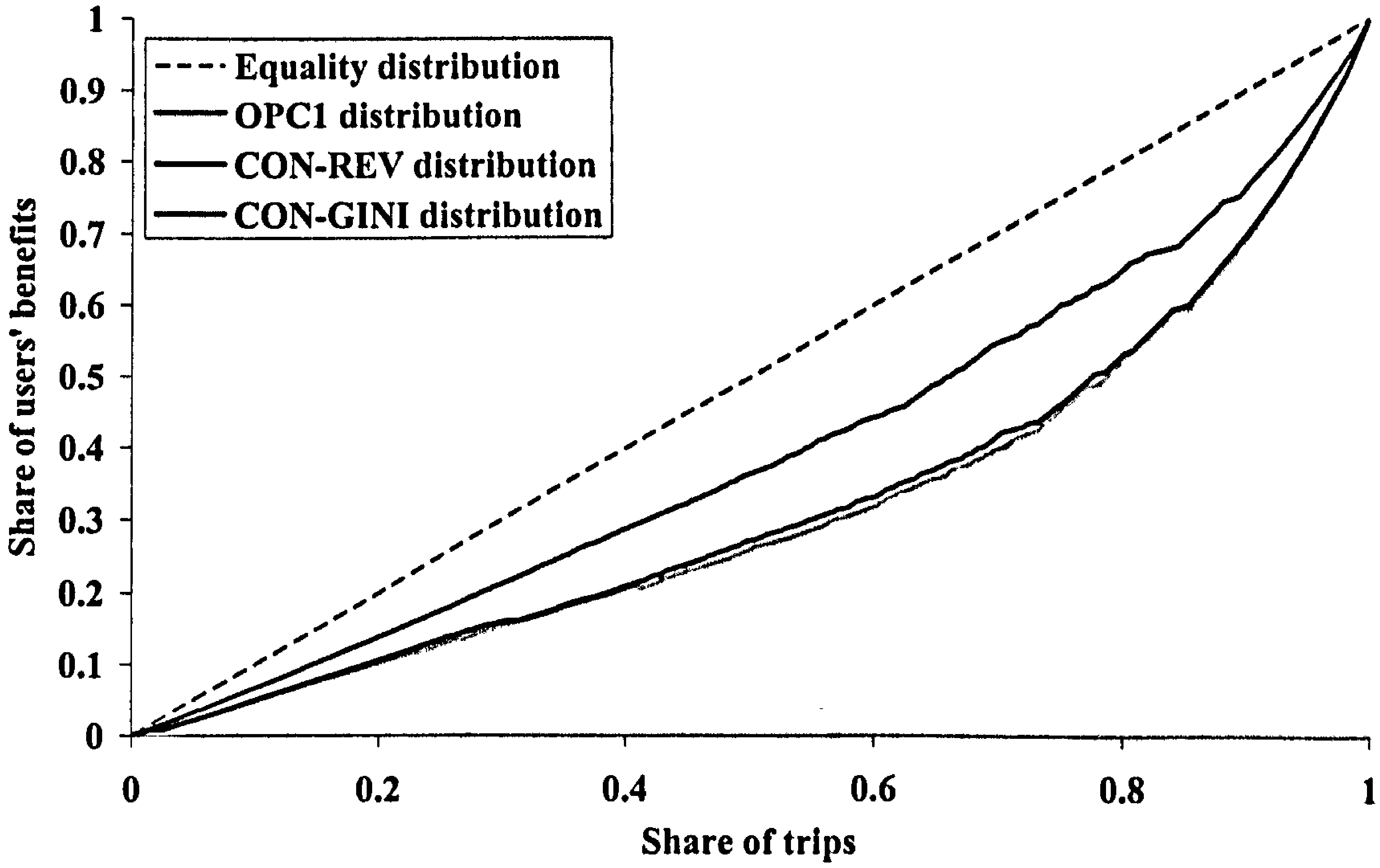


Figure 7-6 Users' benefits distribution for different charging cordons

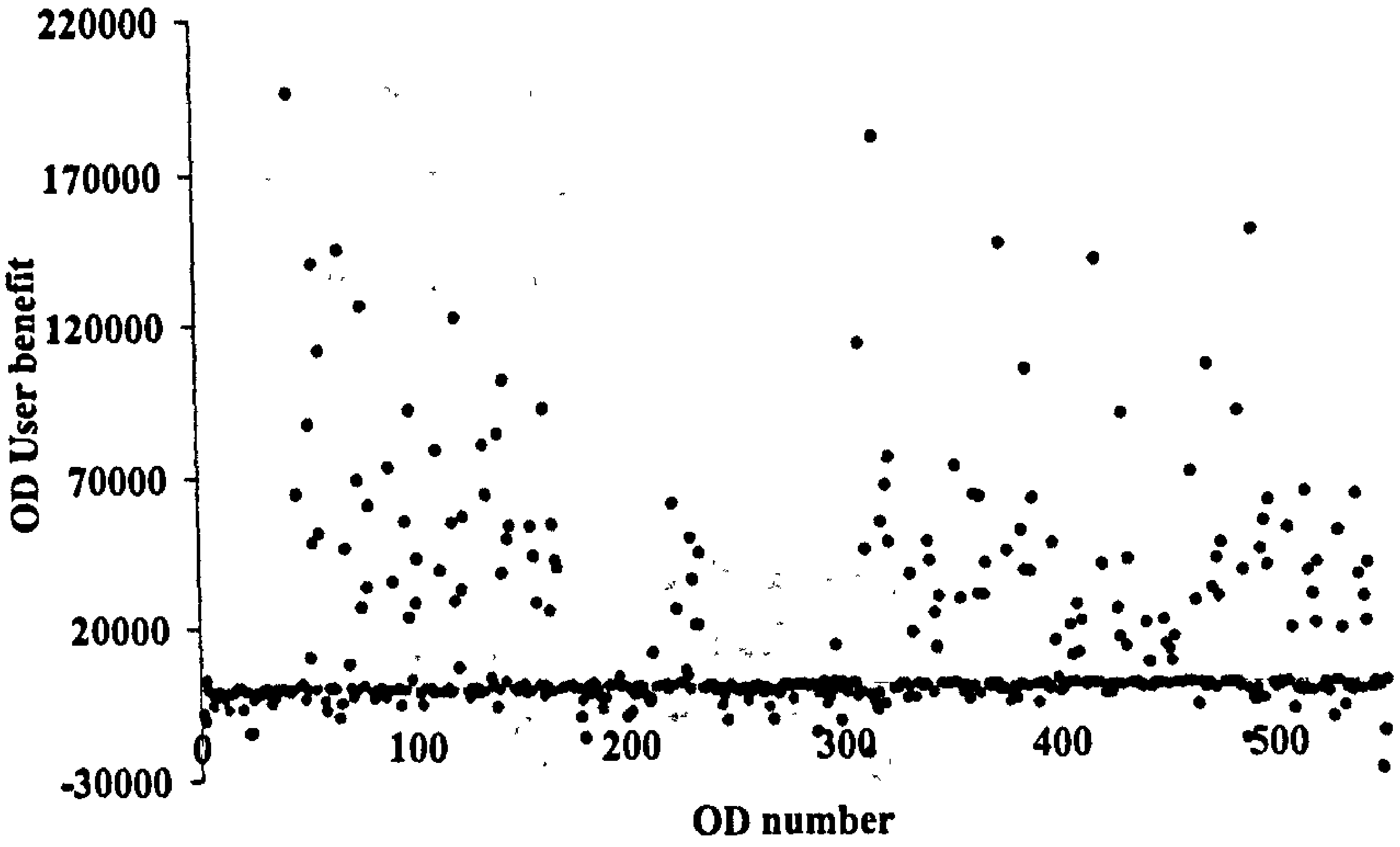


Figure 7-7 User benefits by O-D pairs for OPC1

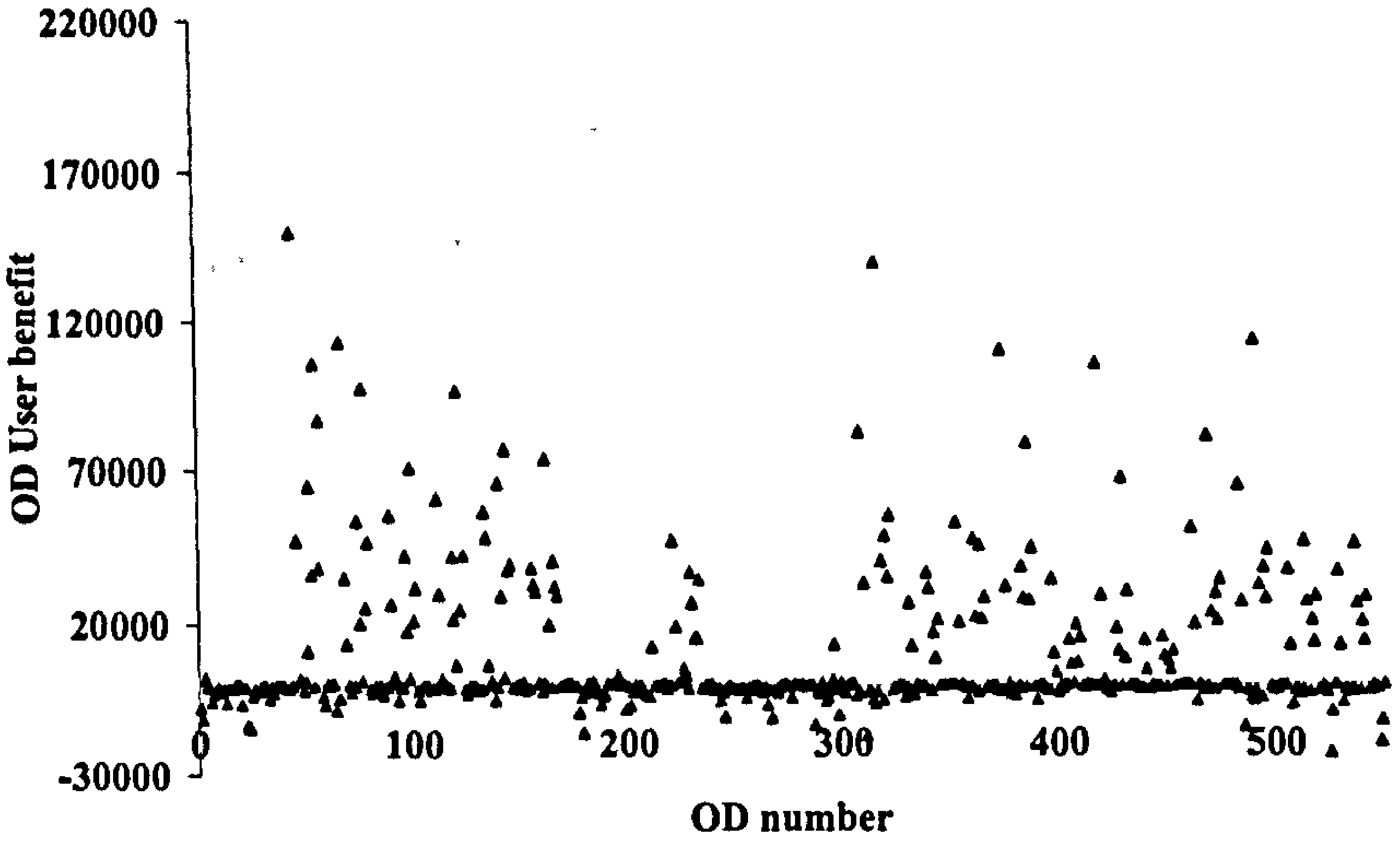


Figure 7-8 User benefits by O-D pairs for CON-REV

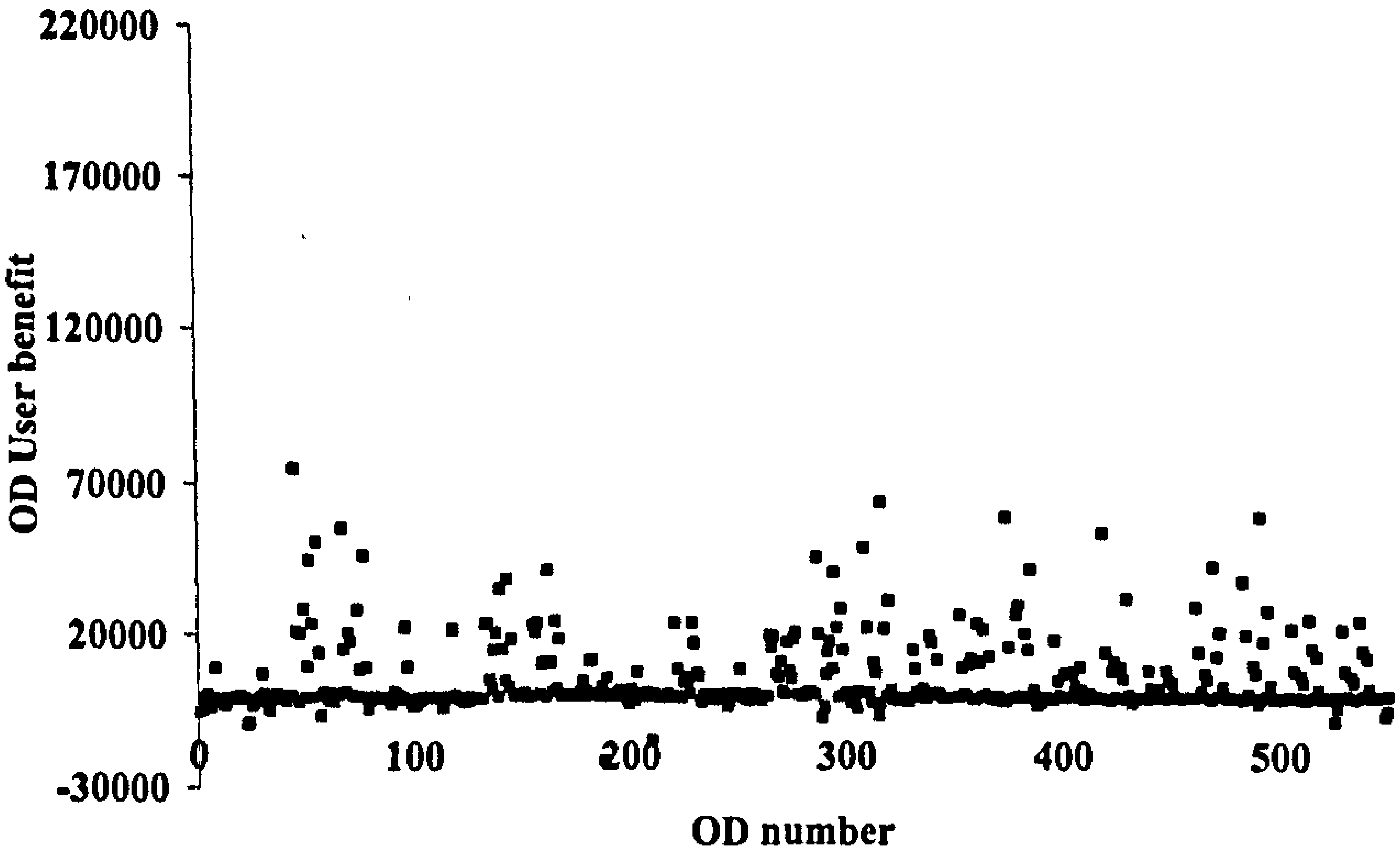


Figure 7-9 User benefits by O-D pairs for CON-GINI

Interestingly, the solution found for the third problem (constraint on total travel time) is the same cordon as the CON-REV but with a toll level of £3.00. The solution for the third problem is referred to as CON-TTTIME. The net benefit for this cordon is around £3.75k per hour which is 40% lower than the net benefit from the CON-REV. The total travel time for the CON-TTIME is around 48910 PCU-Hr/hr which is, as expected, lower than the constraint set of 49500 PCU-Hr/hr.

Figure 7-10 shows the solution for the fourth test (with constraint on the total travel distance). This cordon is referred to as CON-TTDIS. The optimal toll level found for this solution is £2. The total travel distance from this solution is 1687498 PCU-Km/hr, which satisfies the constraint set (TTdis ≤1690000 PCU-Km/hr). The number of tolled point for this cordon is 17 and as shown in the figure the cordon covers a wider area of the network compared to the other cordons discussed earlier. The net benefit of CON-TTDIS is around £4.5k per hour which is lower than the net benefit from the OPC1 by around 38%. Of course, this is traded off with the lower level of total travel distance (the total travel distance for the OPC1 is 1720864 PCU-Km/hr).

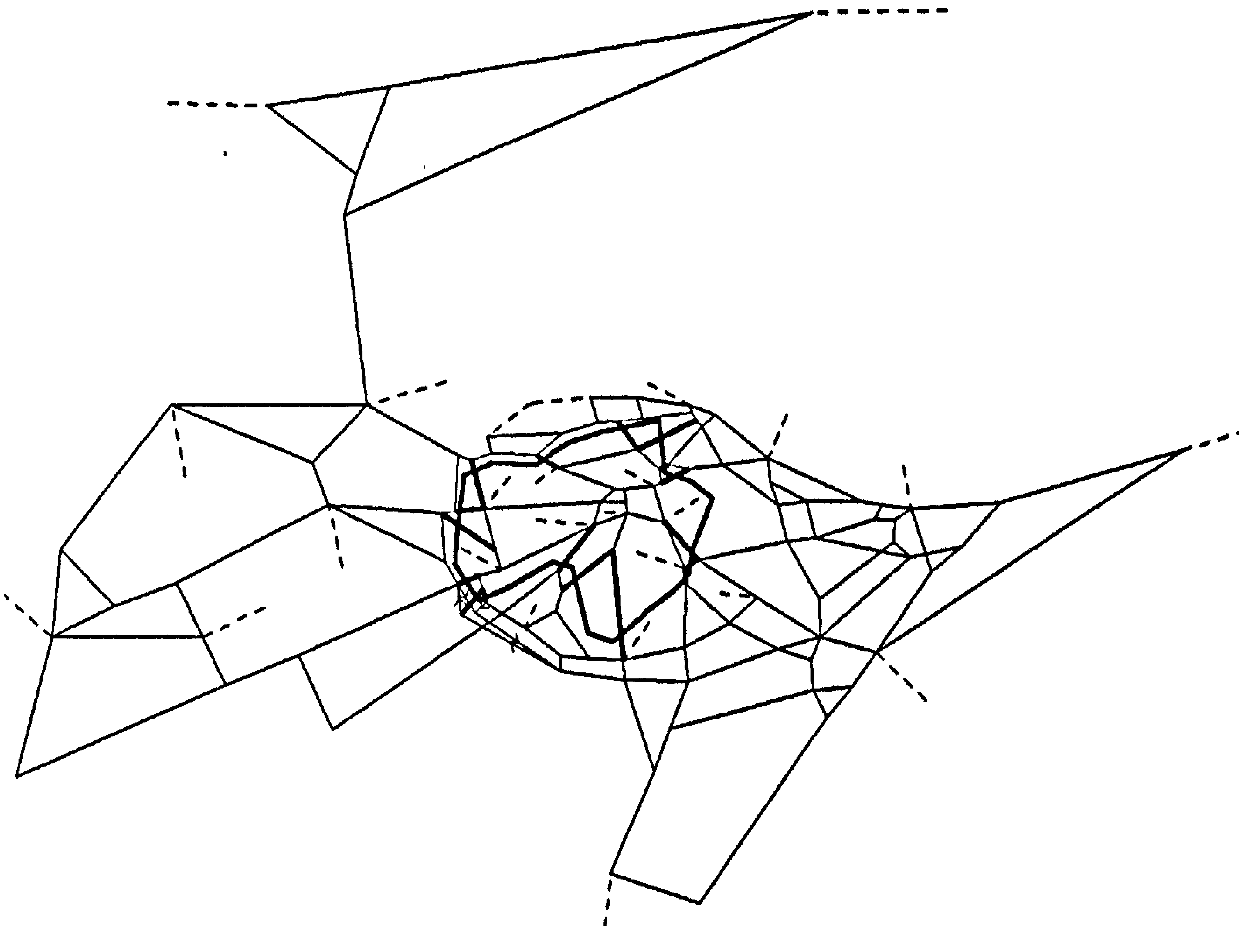


Figure 7-10 Optimal cordon location with constraint on total travel distance (CON-TTDIS) with toll level of £2.00

The last test in this section is to include both constraints on the revenue and equity impact into the cordon design. The first attempt was conducted by including the same constraints as used in the last two tests (net revenue ≥£50k and *Gini* coefficient ≤ 0.30). Interestingly, the outcome design with these constraints yields a negative net benefit. This implied that the constraints imposed are too restricted for defining a scheme with positive net benefit. As a result, the revenue constraints were slightly relaxed by decreasing the minimum level of net revenues required from £50k to £45k. The constraint on *Gini* coefficient was unchanged. With these modified design constraints, CON-GAAS can find a solution with a positive net benefit. Figure 7-11 shows the location of the solution cordon, named CON-REV-GINI.

The key performance of CON-REV-GIN is presented is Table 7-3 above. As expected, the revenue generated (£48k) and *Gini* coefficient (0.29) both satisfy the constraint levels set. The net benefit of the scheme, however, is significantly worse than those of CON-REV and CON-GINI. Comparing with the benchmark of OPC1 in Table 7-3, the net benefit of CON-REV-GINI is about 40% lower. There is a clear trade-off between including both constraints into the design and the level of net benefit of the scheme.

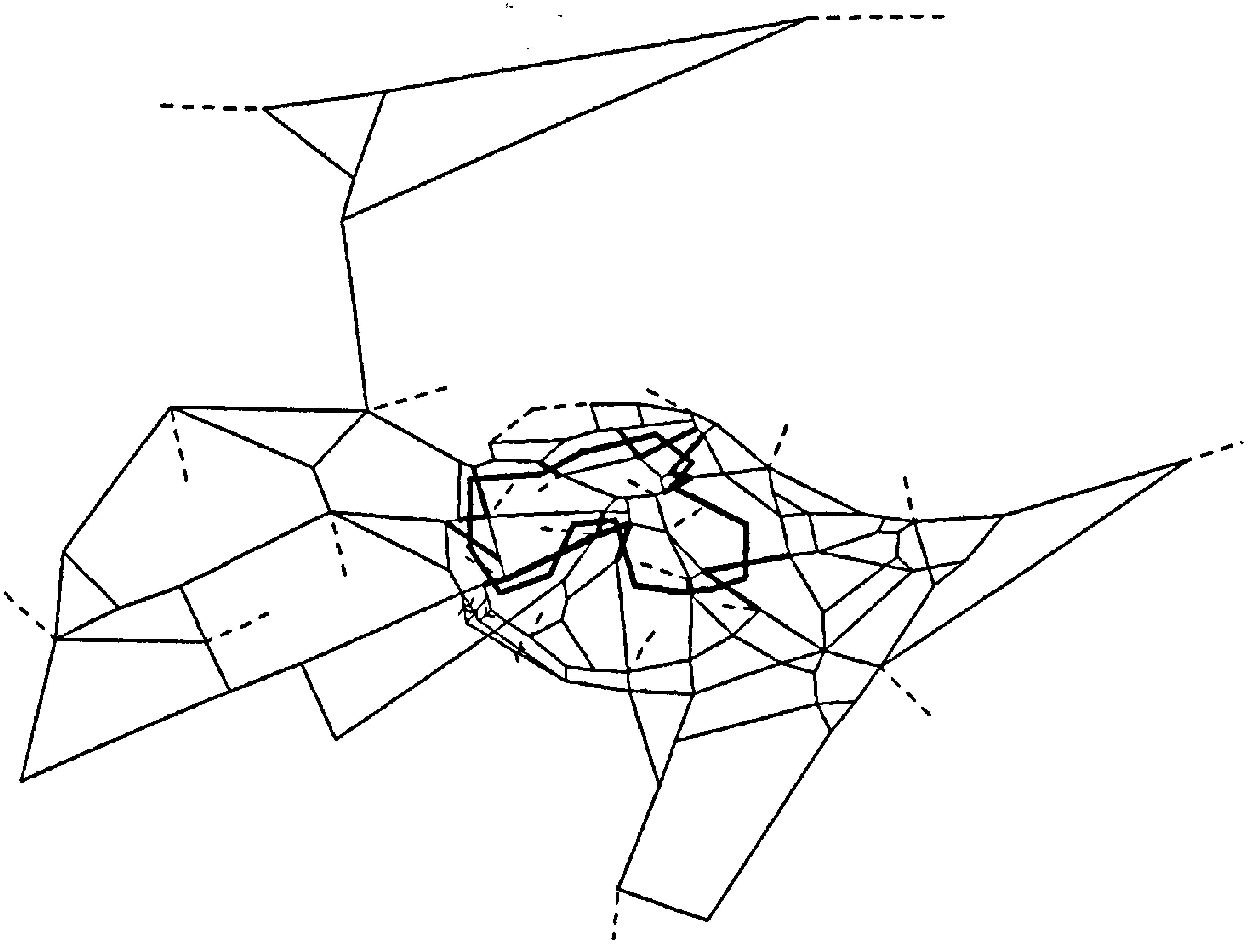


Figure 7-11 Optimal cordon location with constraints on equity impact and net revenues (CON-GINI-REV)

7.5 SUMMARY

The GA-AS algorithm developed in the previous chapter for designing a charging cordon was extended in this chapter to include several constraints into the design. This development followed the discovery in Chapter 2 and Chapter 3 that the practitioner commonly considers different kind of constraints in deciding upon the structure and location of the road pricing scheme.

The chapter reviewed the development of the approach for allowing the GA to deal with constraints. Three categories of methods were discussed including the constraint relaxation approach, the interior feasible space search, and the hybrid method.

Different methods are appropriate for different kind of problem. For instance, the special chromosome encoding/decoding based on the branch-tree strategy discussed in the previous chapter falls into the category of interior feasible space search (dealing with the constraint on the topology of toll points) that is suitable for constraints related to the structure of the solution. However, this approach may not be appropriate for the constraints related to the outcome of the solution since it is difficult to construct a searching mechanism that can ensure the feasibility of the solutions. Thus, in this chapter the approach based on the constraint relaxation method is adopted. In particular the self-adaptive penalty based algorithm is implemented with GA-AS for dealing with the constraints on the cordon design; the method is named CON-GAAS.

There are also other possible penalty based approaches including the static and dynamic penalty methods. These three different penalty based methods were tested with the network of Edinburgh city to maximise the net benefit subject to the minimum level of net revenue generated. The results show the insightful information about the operation of different penalty methods. Using the static penalty based method resulted in very flat trends of the average net benefit and net revenue. This is reasonable given the underlying constant function of the static penalty. Intuitively, the behaviour of the dynamic and self-adaptive penalty methods was different from the static one. The intention of these methods is to allocate more effort to optimising the main objective function in early generations (using a low penalty term), and then gradually increases the penalty as the number of generation increases. As a result, the average net revenues in both the dynamic and self-adaptive penalty methods increased with the number of generations.

A small difference between the dynamic and self-adaptive penalty approach was also found from the results. The self-adaptive penalty actually uses the feedback from the search algorithm (the average net benefit in the previous generation) to adjust the penalty level. Hence, the increasing trend of the net revenue from the self-adaptive penalty method was found to be more rapid than the trend from the dynamic penalty method. In summary, different penalty based methods behave differently due to their underlying formulation. All three methods tested in this chapter successfully found the same final solution. It is generally difficult to compare the performance of different methods. In this chapter, the main method adopted is the self-adaptive penalty method.

CON-GAAS was also tested with different problems again with the network of Edinburgh. The main objective function in all tests was to maximise the net benefit. Five different tests were conducted. The first test was associated with the constraint on the minimum level of net revenue required. The second test included the constraint on the equity impact (using the *Gini* coefficient). The third and fourth tests introduced constraints on the maximum level of total travel time and total travel distance respectively. The last test demonstrated the ability of the CON-GAAS to deal with mul\tiple constraints (two constraints on net revenue and equity were imposed). CON-GAAS successfully found solutions for all problems satisfying the constraints set.

In all tests, there were different levels of reduction in the net benefit compared to the net benefit from the optimal cordon without any constraint (OPC1). The constraint affecting net benefit most was the constraint on total travel time (the third test) in which the net benefit was reduced by 48%. The level of trade off will depend on both the level of constraint set and the type of the constraint. Nevertheless, there were clear trade-offs between adding additional constraints to the design and the level of net benefit achieved for all cases tested. This result demonstrates the inevitable trade-off between the practical aspect of the cordon design and its optimality. However, the detailed discussion of the results was not given in this chapter and this will be treated in Chapter 9 which will concentrate on the policy implications of the results.

CHAPTER 8 CHARGING CORDON DESIGN WITH MULTIPLE OBJECTIVES

Some problems are so complex that you have to be highly intelligent and well informed just to be undecided about them. Laurence J. Peter

8.1 INTRODUCTION

Many real world optimisation problems involve multiple objectives. For instance, if one is making a decision on house location, a number of desired criteria can be included in the decision process, e.g. house price, location, or land space.. The assumption of a single objective function made in most of the optimisation problems is thus too subtle. The review in Chapter 2 revealed an interesting development of road pricing as an urban management tool from its origins in economic theory. As an urban management tool, road pricing needs to serve several objectives for maintaining the quality of life and the economic viability of the city. The survey results with several local authorities in Chapter 3 confirm this point; several objectives including reducing congestion, managing city centre, protecting environment, increasing efficiency, and generating revenue were identified as the purposes for adopting road pricing in different cities.

Despite this fact, surprisingly the development of the optimal design of the road pricing scheme reviewed in Chapter 5 mainly concentrates on the case of optimising the single objective of economic efficiency. This chapter aims to develop a method to deal with the multiobjective optimisation problem (MOOP) for a charging cordon design. The problem associated with the charging cordon design with multiple objectives is referred to as MOOP-CD (Multi-Objective Optimisation Problem for Cordon Design).

There are several ways to deal with this problem as discussed later in Section 8.2. The approach based on the generation of *'Pareto solutions'* is adopted as the main approach in this chapter. The method based on the concept of Non-dominated Sorting Genetic Algorithms II (NSGA-II) is integrated with the branch-tree framework for solving the MOOP-CD problem. In solving the MOOP, there are three possible decision paradigms including (i) Priori preference articulation, (ii) Posterior preference articulation, and (iii) Progressive preference articulation. The first approach is probably the most common framework adopted in the transport area, e.g. the utility function based approach and COBA (Bristow and Nellthorp, 2000; Morisugi, 2000; Emberger *et al*, 2003). However, it is sometimes difficult to decide upon the weight or utility without knowing the actual trade-off between the levels of achievenment of different objectives. NSGA-II is adopted to generate a set of non-dominated solutions where the planner can choose the best compromise cordon design from this set of solutions (Posterior preference articulation). In this chapter, a more advanced method that introduces the interaction between the planner and the optimisation routine is also developed: Progressive preference articulation.

There are six further sections. Section 8.2 describes two key elements in making the decision on the multiobjective problem: the preference format, and preference articulation approach. Section 8.3 reviews different existing optimisation algorithms using both classical and evolutionary based methods for the MOOP. Then, Section 8.4 explains the NSGA-II method for solving the MOOP-CD. The method presented in Section 8.4 is based on the concept of posterior preference articulation approach as discussed in Section 8.2. The other alternative and more innovative approach is the progressive preference articulation approach. Section 8.5 explains the approach to modify the NSGA-II to be applicable to the progressive preference articulation decision paradigm. The methods developed are then tested with the network of Edinburgh in Section 8.6. The tests involve the priori, posterior, and progressive preference articulation based approaches. Finally, the chapter is concluded in Section 8.7.

8.2 DECISION APPROACH TO MULTIOBJECTIVE PROBLEM

The formulation of the multiobjective optimisation problem was presented earlier in Section 4.3.2 in Chapter 4. The difficulty in solving this problem is clearly due to the existence of multiple objectives. This section discusses two important components of the decision process for the multiobjective optimisation problem: preference articulation and decision paradigms for the multiobjective problem. Most of the approaches for the multiobjective optimisation problem are based on the fundamental concept of the ‘non-dominated solution' or the 'Pareto solution'. The definition of the 'non-dominated solution' is given below.

Definition 8-1: *Non-dominated solution or Pareto solution.* Given a set of objective functions considered, $\psi_1,\dots,\psi_q$, (assuming maximising all objectives without loss of generality), a solution $x^{(1)}$ is said to dominate the other solution $x^{(2)}$, if both of the conditions below are satisfied:

(i) The solution $x^{(1)}$ is no worse than $x^{(2)}$ in all objectives, or $\psi_q\left(x^{(1)}\right)\geq\psi_q\left(x^{(2)}\right)$ for all q, and the solution $x^{(1)}$ is strictly better than $x^{(2)}$ in at least one objective, or $\psi_q\left(x^{(1)}\right)>\psi_q\left(x^{(2)}\right)$ for at least one q.

(ii) If any of these two conditions is violated, the solution $x^{(1)}$ does not dominate the solution $x^{(2)}$. If $x^{(1)}$ dominates $x^{(2)}$, we can write $x^{(1)} \succ x^{(2)}$.

Definition 8-2: *Pareto front.* Pareto front is a set of non-dominated solutions.

8.2.1 Preference articulation

The most important process in multiobjective optimisation regardless of the actual optimisation method adopted is to define some form of preference amongst the different objectives involved in the decision. The decision maker, i.e. transport planner or politician, will be presented with the list of possible solutions, e.g. charging cordons, which are equally good in the sense of Pareto optimum (see Definition 8-1 above). The notation of Pareto-optimality is only a first step toward solving a multiobjective problem. In order to select suitable compromise solution from all non-dominated solutions, the decision maker needs to apply his or her preference either explicitly or implicitly to the selection process. Various forms of preference can be defined including:

(i) *Weighting coefficient and utility based approach*: This is probably the main approaches practically adopted in the project evaluation in transport area format (Lee Jr., 2000; Morisugi, 2000) using the cost benefit analysis as the monetary evaluation method for assessing the net benefits of projects (Layard, 1997). For example, in the U.K. under the system of cost benefit analysis, 85% of the appraisal weighting was given to projected time savings for drivers with the remaining 15% of benefits being accounted for by reduced vehicle operating costs and improved safety (House of Commons, 1990). However, the weakness of this paradigm is the potential difficulty for the decision maker to express his or her preference between the objectives without being presented with the actual possible non-dominated solutions (the process with priori articulation of preference). This problem is also well addressed in the context of transport project appraisal (Sayers *et al*, 2003). A wide range of the experimental economic methods (e.g. stated preference method) is available as the tool to overcome the preference assignment problem. If the decision maker is presented with the real set of non-dominated solutions, then he or she does not need to express the weight coefficient or utility for each objective explicitly.

(ii) *Priorities exposition*: Rather than specifying a clear trade-off between the values for different objective functions, the decision maker can express his/her preference by determining in which order objectives are to be optimised, according to their importance.

(iii) *Expressing goals and targets*: The decision maker can express his/her aspirations for different objectives. Goals or targets are considered easier to set than weights and priorities. In transport, there has been a trend toward using the goal and target based approach. With its close mapping with the real performance of the design, it is recommended in the U.K. that the local authorities should set up their transport plans with a clear set of targets(DETR, 2000). Emberger *et al* (2003) recently proposed a method to optimise city strategic transport policy with a target-based approach as contrast to the conventional cost-benefit analysis approach.

(iv) *Linguistic ranking*: This is probably the most recent approach. The psychological aspect causes some difficulty for human (as the decision maker) to precisely express his/her preference. The linguistic ranking method is developed to allow the decision maker to express his/her preferences in a more fuzzy form (Chen *et al*, 1992). For instance, the preference for different objectives can be expressed verbally, e.g. more important, less important, equally important, no preference. This approach has been recently attracting some attention from the research community. For example, Cvetkovic and

Parmee (2002) introduced the concept of fuzzy preference with linguistic ranking with GA for the multiobjective optimisation problem.

8.2.2 Decision paradigms for the multi-objective problem

The process of identifying the preference (or preference articulation) can be conducted in several ways. The stage of the preference articulation naturally depends on how the optimisation method is applied to solve the problem. Three possible major classes of the multiobjective decision paradigms can be defined (Hwang and Masud, 1979):

(i) *Priori articulation of preferences:* In this paradigm, the decision maker expresses preferences toward different objectives prior to the optimisation process. Then, this set of preferences is used to combine all objectives into a single utility term reducing the problem to a single objective optimisation problem. From the computational point of view, it is easier to reduce the multiobjective problem to a single objective problem and then solve it. With the predefined weights for each objective function, the weight-summed objective function can be created. Then, any kind of optimisation methods for a single objective optimisation problem can be applied to the problem. However, if the preference is expressed in the form of priority, solving the optimisation problem is still a difficult process; and as discussed earlier it is often difficult to determine the weight without seeing the possible solutions.

(ii) *Posterior articulation of preferences:* In this paradigm, the optimisation method is applied to the problem to define the set of non-dominated solutions. Once the decision maker is presented with this set of non-dominated solutions, she then expresses her preferences on the choices, resulting in the most preferred compromised solution.

(iii) *Progressive articulation of preferences:* This approach is referred to as the learning approach by Vreeker et al (2002); it is based on a sequential (interactive or cyclical) articulation of the decision maker's views on the best compromise solution. From the operational context, the decision making and optimisation process occur in interleaved steps. At each step, the decision maker supplies partial preference information based on the current non-dominated solutions to the optimisation algorithm. Then, the optimisation process, in turn, generates a new set of non-dominated solutions situated nearer to the actual region of interest of the decision maker (in terms of the compromise between different objectives).

The first type of preference articulation process is arguably the most popular method adopted for planning and evaluation purposes. This is due to its simplicity and robustness. However, setting the preferences (e.g. weight, priority, or goal) is often difficult if the decision maker does not have any information about the true trade-off between different objectives, e.g. the transport planner does not know exactly the list of possible trade-offs between the congestion reduction and revenues generated from the road pricing scheme.

On the other hand, the posterior preference articulation approach presents the decision maker with the set of Pareto solutions so that she can make decisions on her preferences with information on the real trade-offs. This approach has not been widely adopted. This may be due to the slow development of a good optimisation method for defining high quality non-dominated solutions. The development of evolution optimisation algorithm in the last decade does increase the possibility of using this approach for the decision making process (this approach is adopted in this chapter and presented later in Section 8.4).

However, despite the fast development of the evolutionary optimisation, it is still extremely expensive to construct the full description of the trade-off surface. The approach of progressive preference articulation has the potential advantage of reducing computational effort required by concentrating the computational effort on the region of interest from the decision maker's point of view (which is received interactively). On the other hand, implementing the progressive preference articulation method is much more complicated compared to the first two approaches mentioned earlier. The main difficulty is the interface between the decision maker and the optimiser. In this chapter, the progressive preference articulation method is also tested as explained later in Section 8.5.

8.3 REVIEW OF GA APPROACHES FOR MULTIOBJECTIVE OPTIMISATION PROBLEM

This section reviews classical and evolutionary optimisation methods for tackling the MOOP.

8.3.1 Classical methods

There are two main classes of classic optimisation method for solving the MOOP: the utility aggregation approach and the goal or target based approach.

Utility aggregation approach

The first strategy aims to reduce the multiobjective functions to a single objective function through some kind of utility aggregation method. The weighted sum method is probably the simplest and most favourite method adopted in this category. As the name suggests, the method aggregates set of objectives into a single objective by multiplying each objective with a user pre-defined weight.

The main theorem adopted in this approach is that given a set of weights, the solution to the weighted sum problem will definitely be one of the non-dominated solutions on the Pareto front with a convex feasible space and objective function (Chankong and Haimes, 1983). Friesz and Harker (1983) adopted the weighted sum approach for solving the muticriteria network design problem using the Hooke and Jeeves method for the transformed single level non-linear optimisation problem. Friesz *et al* (1993a) similarly adopted the weighted sum approach but with the simulated annealing method to solve the multiobjective network design problem. However, the weighted sum approach faces several drawbacks. Firstly, the method cannot deal with the non-convex problem, i.e. the algorithm is not guaranteed to find a true Pareto front. Secondly, a uniformly distributed set of weight vectors need not find a uniform distributed set of Pareto solutions and different weight vectors need not necessarily lead to different Pareto solutions.

To deal with the weighting problem, a number of devised weighting schemes have been proposed. These include various types of weighted metric methods, e.g. weighted *Tchebycheff*, Benson's method , or the rotated weighted metric method (See Miettinen, 1999 for more details of these methods). Regardless of the weight form, the common philosophy of weighted metric method, in contrast to the normal weighted sum approach, is the provision of a reference point (representing the ideal solution) in which each objective will be measured by the relative metric distance to this ideal point. Then, a weighted sum of the metric distances of all objectives can be calculated. With a good location of the ideal solution, the weighted metric approach guarantees finding each and every Pareto solution (Miettinen, 1999).

An alternative approach that allows a more general form of aggregated objective is the value function method (Keeney and Raiffa, 1976). With this approach, any form of utility function relating all objectives can be adopted to define the aggregated objective function provided that the utility function satisfies the strongly decreasing condition (Rosenthal, 1985). Both weighted sum and weighted metric approach can be considered as a special case of the value function method.

Goal or target based approach

A rather different strategy to the utility aggregation approach is the target or goal based approach. The main mechanism of this approach is the conversion of the objectives to a set of constraints through the set of goals or targets. The decision maker in this problem will have to specify a target or goal (aspiration level) for each objective. Then, two approaches for converting the objective to a constraint can be adopted including the ϵ-constraint method (Haimes *et al*, 1971) and the goal programming method (Ignizio, 1976). With the ϵ-constraint method, only one objective will be left as the main objective function and the other objectives will be converted to a set of hard constraints (i.e. $\psi_i(x) \le b_i$ where ψ_i is the objective function i and b_i is the goal/target of this objective function). The hard constraint implies that in the final solution the constraint must be definitely satisfied.

On the other hand, if the goal programming approach is adopted, all objective functions are converted to soft constraints (i.e. $\psi_i(x)+w_i\cdot\lambda\le b_i$ where w_i is the normalised weight factor for objective i that the designer must define beforehand, and λ is the dummy variable used to formulate the main objective function). The main objective function of the goal programming problem is to minimise the un-achievement level:

$$\begin{aligned}&\min_{\lambda,x}\lambda\\&s.t.\\&\psi_i(x)+w_i\cdot\lambda\le b_i\quad\forall i\end{aligned}$$

Therefore, the constraint can be seen as the soft constraint that may not be necessarily satisfied by the final solution.

These two different classical optimisation approaches to the MOOP are associated with different types of preference expressions as explained in 8.2.1. The aggregated utility function approach can be seen as the optimisation mechanism for the utility based preference whereas the goal/target approach is suitable for the other two preference formats (i.e. goal and priority).

Despite the possible application of the classical optimisation method to the MOOP, there still exist significant differences in the nature of solving MOOP and a normal single objective optimisation problem. For instance, the aim of a single objective optimisation is to find a single solution whereas solving the MOOP naturally involves generating a number of final solutions (e.g. Pareto solutions). The restriction of the classical optimisation methods does not allow a natural treatment with the MOOP. Recently, the evolutionary optimisation algorithm (EA) has been the major interest for the researchers working in the area of MOOP. EA mainly operates on the population basis. This characteristic fits in naturally with the task of finding a set of solutions rather than a single solution for the MOOP. In addition, the flexibility of the EA allows a more arbitrary form of the preference as well as of the objective functions and constraints. The next section reviews the development and current state of the art of the evolutionary based algorithm for the MOOP.

8.3.2 Evolutionary based methods

As mentioned, the population based search nature of GA makes the approach very suitable for solving the MOOP. Since the realisation of its potential, many researches have been investigating and developing efficient GA based methods for the MOOP. Many utility aggregation approaches explained in Section 8.2.1 can also be used with the GA in which the multiobjective problem can be reduced to a single objective problem and any GA methods can be applied to the problem directly. However, this way of using GA does not exploit the whole possible potential of GA as the searching mechanism for the MOOP. Three main research questions challenging the GA research community includes (i) the optimality of the Pareto solutions, (ii) the efficiency of the search algorithm, and (iii) requirement of a uniform spread solutions(Zitzler and Thiele, 1999; Deb, 2001). In brief, the ideal GA based method for the MOOP must be able to identify the true Pareto front with well spread Pareto solutions in the most efficient process. Two main categories of the methods are discussed: non-Pareto and Pareto based methods. The non-Pareto based approach does not use the comparison of the domination of the chromosomes to determine their survival; but the Pareto-based method does.

Non-Pareto based methods

Early development has concentrated on devising some kind of framework to maintain the original framework of GA. This group of approaches is referred to as the non-Pareto approach. One common approach adopted in this category is to partition the evolution process for each objective; hence the normal GA selection can be used for each sub-population and then introduce some possible interaction can then be introduced between different sub-population groups. The approaches in this group rely on the belief that a combination of two chromosomes that are fit for each objective will lead to a chromosome that is fit for both objectives.

The first implementation of this kind of approach is by Schaffer (1985). Schaffer implemented the vector evaluated genetic algorithm (VEGA) for the MOOP. The modification of this approach and the normal GA lies in the selection process. A number of sub-populations will be generated by the selection process in which each sub-population is associated with the selection process for each objective. Then, the chromosomes from the whole sub-population will be shuffled and the normal GA process will proceed.

The approach represents an early attempt in tackling the MOOP with GA with a simple modification to the original GA. However, the main weakness discovered later on is that the algorithm may generate a number of super-fit chromosomes for each objective. The chromosome with some balance compromise between different objectives may not be survived during the selection process.

Other approaches proposed that fit in with this category include the lexicographic ordering approach (Fourman, 1985), game theory approach (Jacques *et al*, 1997), and gender labelling approach (Allenson, 1992). Briefly, the lexicographic method requires the user to specify the order of importance of the objectives. Then, GA is used to find the optimal solution for the first objective producing the highest achievable level of the first objective. Then, a second problem is constructed considering the second objective as the main objective with the constraint that the first objective must be maintained at the level achieved in the previous optimisation. The second problem is then solved by applying GA. The process carries on until the last objective is considered. This method is relatively simple compared to the utility aggregation approach and VEGA. However, the method can be classed as the utility aggregation approach. This is because the decision maker is asked to express the preference explicitly prior to the optimisation process (in the form of lexicographic ordering). Hence the algorithm is unable to generate the whole range of non-dominated solutions.

For the game theory based approach, the concept of Nash's equilibrium is adopted where a number of sub-populations are created who represent the players wishing to maximise different objectives. Each sub-population is evolved separately including the selection process (unlike VEGA approach). Then, the interaction between different groups occurs through the migration process where the best chromosome in each sub-population will be transferred to other sub-populations. The process is repeated for as many generations as needed, until the Nash equilibrium is reached. This method is claimed to be very efficient in finding a non-dominated solution. However, the drawback is that the method may only be able to identify only one non-dominated solution, rather than the whole range of solutions on the Pareto front.

The gender labelling approach is similar to VEGA where a number of sub-populations associated with different objectives are created. Before shuffling the whole sub-populations together, each chromosome is randomly assigned its gender (male or female) and the approach ensures the equal assignment between male and female populations. Then, the selection process carries on as VEGA. In the mating process, only the chromosomes from different genders can be mated for the crossover process. Lis and Eiben (1996) propose the idea of multi-parent crossover (referred to as panmictic reproduction) for the case with more than two objectives (in which case there are more than two genders). The mating restriction in this approach may introduce some computational burden when the number of objectives increases (impling more genders), since the algorithm needs to exhaustively find the right match between different genders and the size of the population must be relatively large to produce diverse children for the next generation.

Pareto based methods

As noted, the development of the algorithm in the group of non-Pareto based aims to reduce the problem to design an approach to apply the normal GA operation to the MOOP. The main drawbacks of these approaches are the possible lack of diversity of the non-dominated solutions generated and the efficiency of the method. The second wave of the search for a good GA approach for the MOOP has moved the centre of interest into introducing the concept of non-dominated solutions inside the GA directly. This class of methods is referred to as the Pareto-based method.

In fact, Goldberg (1989) proposed in his book using Pareto definition to represent the single metric fitness for the chromosome to give a higher chance of survival for a non-dominated chromosome. Nevertheless, there was no significant work or development after the suggestion (Deb, 2001). Goldberg's suggestion also includes a *niching* strategy amongst non-dominated solutions to ensure a good spread of the solutions on the Pareto front. This idea aimed mainly to resolve the problem experienced in VEGA.

Realising the possible advantage of using Pareto based fitness assignment, at least three independent methods have been proposed including Multiple Objective GAs –MOGAs-(Fonseca and Fleming, 1993), niche Pareto GAs –NPGAs-(Horn and Nafpliotis, 1993), and non-dominated sorting GAs -NSGAs-(Srinivas and Deb, 1994). These three methods used Goldberg's suggestion (i) the fitness of a solution was assigned using the extent of its domination in the population, (ii) the diversity among solutions was preserved using the niching strategy.

The slight difference between these three methods is the fitness assignment. For MOGAs, the fitness for each chromosome is set to equal to the number of chromosomes dominating that solution (lower fitness is better in this case). For NPGAs, the tournament selection is adopted where a sub-population is randomly picked from the whole population. The fitness of a chromosome is assigned according to the number of chromosomes in the sub-population that dominates this chromosome. For NSGAs, the chromosomes are initially classified into a number of non-dominated fronts. Then, each chromosome on front *j* is assigned with the fitness of *j* (the lower the better).

These algorithms represent early attempts at introducing directly the idea of non-dominated solutions inside the GA process. However, various numerical experiments showed the lack of convergence of these algorithms to the true Pareto front since an operator for preserving elitism chromosomes was missing (Laumanns *et al*, 2002). Various methods exploiting the elitism strategy have thus been proposed including Strength Pareto Evolutionary Algorithm –SPEA-(Zitzler and Thiele, 1999), Pareto Archived Evolutionary Strategy –PAES-(Knowles and Corne, 2000), and Non-dominated Sorting GAs II -NSGA-II-(Deb *et al*, 2000).

The key development in this family of algorithms is the introduction of the concept of an archive. The archive is used to contain the list of the non-dominated solutions found during the process. The off-springs from the evolution process in each generation are always compared with the archived solutions. If the off-spring is not dominated by any of the archived solutions, it will be preserved and allowed to pass on its genes to the next generation. This concept is adopted in all elitism based algorithms mentioned above with some slight differences in the actual archive updating process.

Some new approaches to increase the diversity of the solutions on the Pareto front were also proposed. Zitzler and Thiele (1999) proposed a clustering algorithm. In SPEA, the combined non-dominated solutions from parent and off-spring are created and the size of this combined population may exceed the size of the population for the next generation. Each chromosome is sequentially assigned to one of the clusters by its distance to the nearest chromosomes (the number of clusters is equal to the number of populations required for the next generation). Then, the chromosome with the minimum average distance from other solutions in each cluster is retained and all other solutions are deleted. The other proposed diversity preservation strategy is the crowding distance method adopted in NSGA-II (Deb *et al*, 2000). The method estimates the density of solutions surrounding a particular chromosome *i* by taking the average distance of two chromosomes on either side of the chromosome *i* along each objective.

Clearly, there are many possible ways to apply GA or EA to the MOOP. Different methods operate in different ways and may perform differently from one problem to another. Some comparisons on the performances of different methods either qualitatively or quantitatively have been conducted but the common consensus on which method is superior to the others has not been reached so far. In this thesis, the method of NSGA-II is chosen as the optimisation algorithm for the MOOP-CD. The next section introduces the detail of the NSGA-II for posterior preference approach. The method is then extended in Section 8.5 to deal with the progressive preference articulation problem.

8.4 NSGA-II FOR OPTIMAL CORDON DESIGN: POSTERIOR PREFERENCE ARTICULATION APPROACH

8.4.1 Overview of the method

As mentioned in the introduction, the approach to the MOOP in this paper is to seek a set *non-dominated solution* (or *'Pareto solution'*). The method proposed in this section is the Posterior preference articulation based method. The main task of the optimiser is to generate the set of non-dominated solutions (a set of different charging cordon designs). With the set of solutions on the Pareto front, the planner can then choose the best compromise cordon design. The important issue from the optimisation point of view is that the algorithm must be able to (i) find or closely approximate the true Pareto front and (ii) find a well-spread set of non-dominated solutions. These two requirements have been the major challenge for the development of the evolutionary based optimisation for the MOOP as discussed earlier in Section 8.3.

The elitist Non-dominated Sorting GA (named as NSGA-II) method proposed by Deb *et al* (2000) is adopted to solve the MOOP in this section. An outline of the algorithm step-by-step is as follows. Initially, a random population (P_0) of toll rings (with toll level) is created. The population is sorted into different non domination levels. This process is to identify different non-dominated fronts from the population. Figure 8-1 shows an example of this process. Chromosomes A, B, and C in Figure 8-1 are at the highest level of non domination. The Pareto front containing these three chromosomes is referred as Front 1. Chromosomes D, E, and F in Figure 8-1 are at the second level of non domination, since the front containing these chromosomes (called Front 2) will become a Pareto front if we remove the front 1 (consisting of A, B, and C). Similarly, Chromosome G, H, and I are at the third level of non domination, since front 1 and front 2 have to be removed so that this front will become a Pareto front.

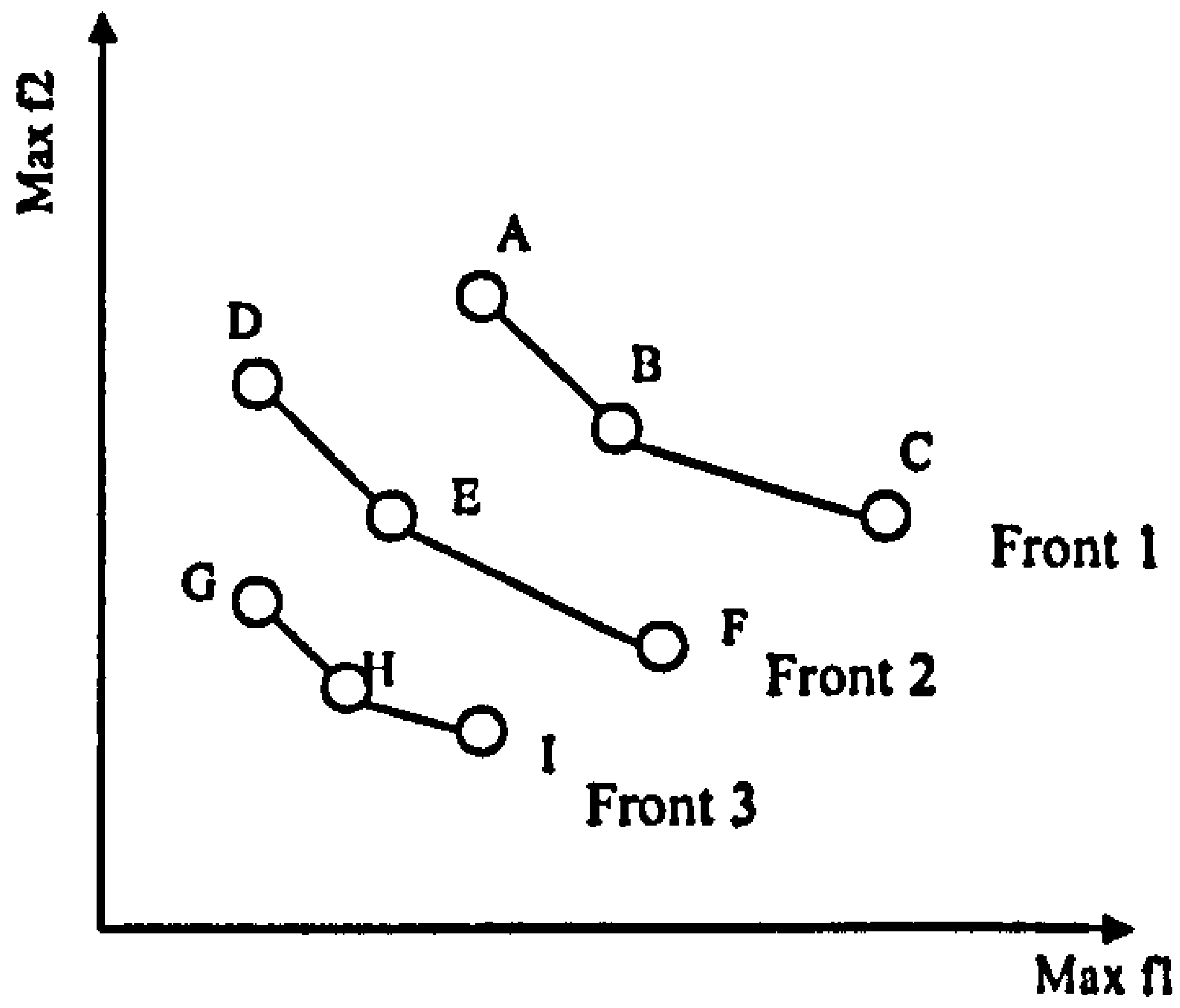


Figure 8-1 Non-dominating level

Each solution is then assigned a fitness equal to its non-domination level (1 is the best level). Thus, in the selection process the lower the fitness value the better the chromosome. Binary tournament selection (with a crowded tournament operator described later), crossover, and mutation operators are used to create offspring population (Q_0) of size N. After this initialisation, the NSGA-II procedure is carried on as follows:

Algorithm NSGA-II

Step 1: Combine parent and offspring populations and create $R_t = P_t \cup Q_t$.

Step 2: Perform a non-dominated sorting to R_t and identify fronts: $F_i, i = 1,2,\ldots, etc.$

Step 3: Set new population $P_{t+1} = \varnothing$. Set a counter = 1. Until $|P_{t+1}| + |F_i| < N$, perform $P_{t+1} = P_{t+1} \cup F_i$ and $i = i+1$, where N is the population number.

Step 4: Perform the Crowding-sort $(F_i, <_c)$ procedure (described in Section8.4.28.4.2) and include the most widely spread $(N - |P_{t+1}|)$ solutions by using the crowding distance value in the sorted F_i to P_{t+1}.

Step 5: Create offspring population Q_{t+1} from P_{t+1} by using the crowded tournament selection (described in Section 8.4.3), crossover, and mutation operators.

Figure 8-3 shows the schematic diagram of the process of NSGA-II. On the left hand-side block, two populations, the parent (P_t) and the offspring (Q_t) populations, are combined to produce the pooled population for the selection process (R_t). All chromosomes are then sorted according to their non-dominated level (described in Section8.4.2). In the middle block, the chromosomes in Front 1 (F_1) and Front 2 (F_2) are totally moved across to the new population set in the right hand-side block. However, as the figure show, the number of the chromosomes in Front 3 (F_3) exceeds the space left for the new population set. Thus, the chromosomes in F_3 are sorted by their crowding distances. Then, a subset of solutions with better crowding distances is included into the new population set. With the new population set (P_{t+1}), the crowded tournament operator (described in Section 8.4.3), crossover and mutation operators (described earlier in 6.4) are applied to P_{t+1} resulting in a new set of offspring for the generation t+1 (Q_{t+1}) and the process iterates as described above until it reaches the maximum number of generation set.

Next, the detail of the key operators for the NSGA-II process is explained including the crowding distance, crowding sort operators, and crowded tournament selection. Note that the chromosome structure, crossover and mutation operators applied in NSGA-II are exactly the same as those explained in Chapter 6 for GA-AS.

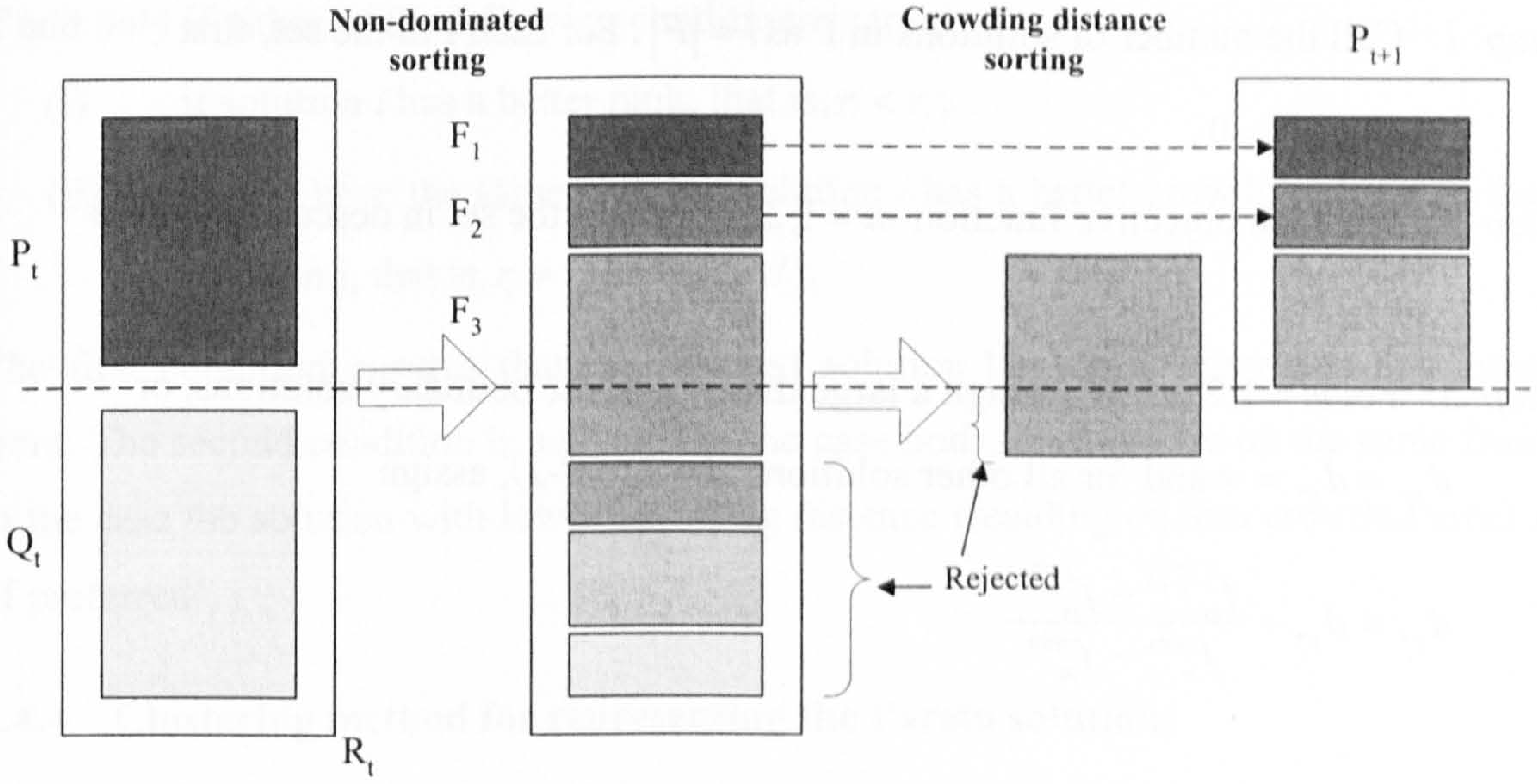


Figure 8-2 Schematic of the NSGA-II algorithm

8.4.2 Crowding distance assignment: Crowding-sort (F_i,$<_c$)

One of the good characteristics of the NSGA-II, as a searching procedure for MOOP, is its ability to define well-spread solutions along the true Pareto front presenting a wide range of options to the decision maker. In NSGA-II, the density of the population in the solution space is adopted as the main indicator for maintaining the spreading of the solutions.

To get an estimate of the density of solutions surrounding a particular solution i in population, we take the average distance of two solutions on either side of solution i along each axis of the objectives. This measure (d_i) represents an estimate of the perimeter of the cuboid formed by using the nearest neighbours as the vertices (called *'crowding distance'*). In Figure 8-3, the crowding distance of the *i-th* solution in its front (marked with a solid circle) is the average side-length of the cuboid (shown by a dashed box). The following algorithm is used to calculate the crowding distance of each point in the set F.

Algorithm: *Crowding-sort*

Step 1: Call the number of solutions in F as $l = |F|$. For each I in the set, first assign $d_i = 0$.

Step 2: For each objective function $m = 1,2,\ldots,M$ sort the set in descending order of f_m.

Step 3: For $m = 1,2,\ldots,M$, assign a large distance to the boundary solutions, or $d_{I_1^m} = d_{I_l^m} = \infty$ and for all other solutions $j = 2$ to *(l-1)*, assign

$$d_{I_j^m} = d_{I_j^m} + \frac{f_m^{(I_{j+1}^m)} - f_m^{(I_{j-1}^m)}}{f_m^{\max} - f_m^{\min}}$$

The index I_j denotes the solution index of the *j-th* member in the sorted list. Thus, for any objective I_1 and I_l denote the lowest and highest objective function values, respectively.

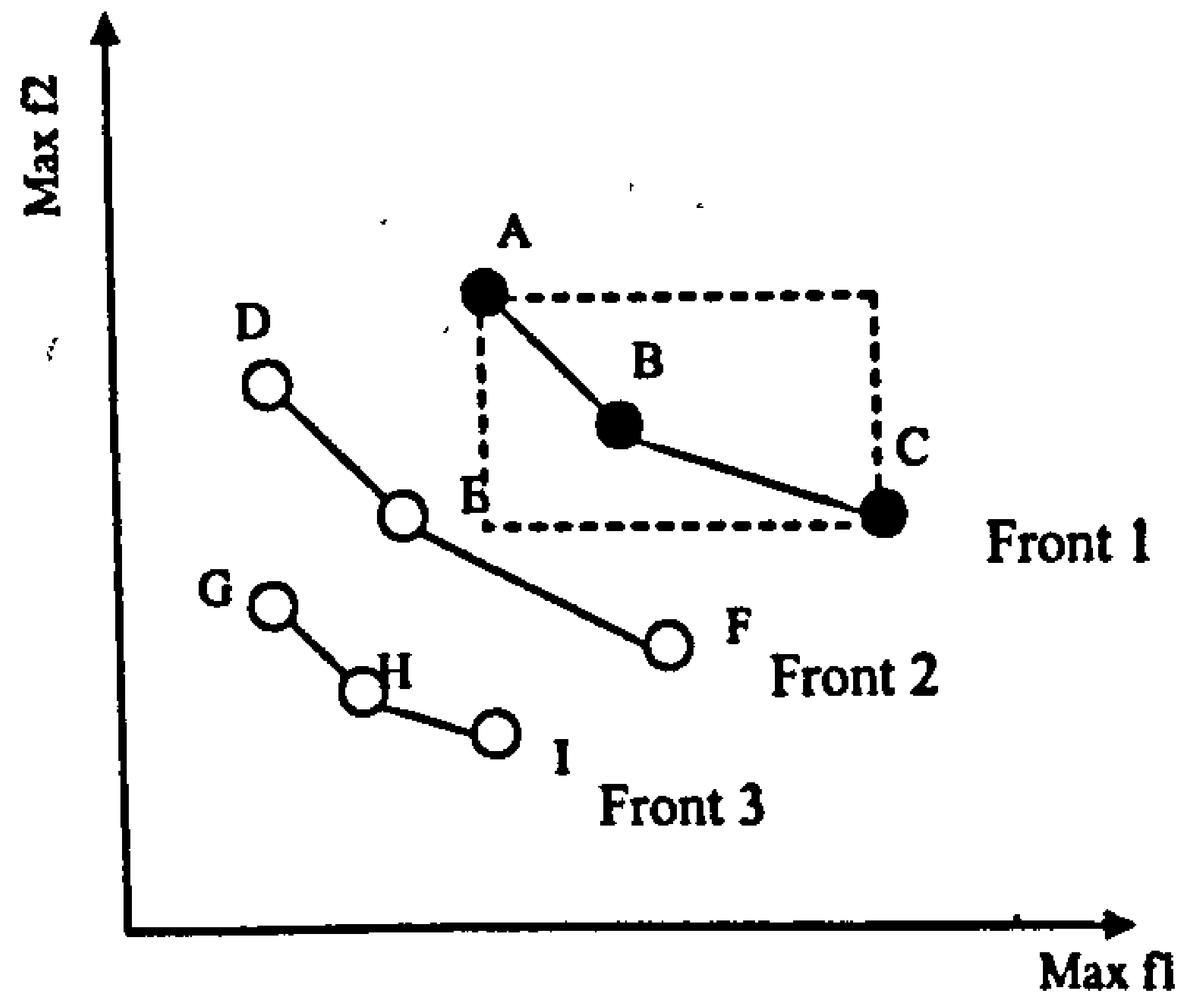


Figure 8-3 Illustration of the crowding distance

8.4.3 Crowding tournament selection operator

The crowded comparison operator $(<_c)$ compares two solutions and returns the winner of the tournament. It assumes that every solution *i* has two attributes:

(i) a non-domination rank (r_i) in the population;

(ii) a local crowding distance (d_i) which is explained above in the population.

In crowded tournament selection, a solution *i* wins a tournament with another solution *j* if and only if either of the following conditions is true:

(i) if solution *i* has a better rank, that is, $r_i < r_j$;

(ii) if they have the same rank but solution *i* has a better crowding distance than solution j, that is, $r_i = r_j$ and $d_i > d_j$.

The first condition ensures that the selected solution lies on a better non-dominated front. The second condition is activated in the case both solutions are on the same front; in the case the solution with lower crowding distance (residing on less crowded area) is of preferred.

8.4.4 Clustering method for representing the Pareto solutions

The NSGA-II method generates a number of non-dominated solutions that will be presented to the DM. The DM then needs to make a posterior preference based decision in order to identify the most suitable/compromise solution from this Pareto set. For the optimal cordon design problem, the planner may face a very high number of possible solutions characterised by different levels of different objectives. The DM may be overloaded with the information and incapable of making a good judgment. In order to simplify the Posterior preference articulation process, the set of non-dominated solutions output from the NSGA-II can be clustered into different groups either based on the levels of objective functions or on the characteristics of the cordon shape.

8.5 NSGA-II WITH PROGRESSIVE PREFERENCE ARTICULATION

8.5.1 Revisit the definition of non-dominated solution

The definition of the non-dominated solution or Pareto solution given earlier (see Definition 8-1) is not the most general definition for the non-dominated solution in the sense that it does not allow different types of preference (e.g. priority or goal). The strategy for progressively articulating the preference from the decision maker explained later in the subsequent section involves using goals and priorities to determine the set of non-dominated solutions. This section extends the definition of the Pareto solution to a more general term in which different kind of preference operators will be allowed.

Let $\psi_i(x)$ $i = 1,\ldots,m$ be the m objectives of the maximisation problem (the definition can be easily converted to the minimisation problem) and ψ is the m-dimensional vector function of some decision variable x. In the general definition of the non-dominated solution, an m dimensional *preference vector* must be specified. The preference vector contains the information about the goal and priority for each objective. The goal considered in this case is in the form of $\psi_i \geq s_i$ where s_i is the goal of the objective i. The definition of the Preference vector is given as follows:

Definition 8-3: *Preference vector.* Preference vector is an m dimensional vector defined as

$\mathbf{s} = [\mathbf{s}_1,\cdots,\mathbf{s}_p] = [(s_{1,1},\cdots,s_{1,m_1}),\cdots,(s_{p,1},\cdots,s_{p,m_p})]$ where p is a positive integer, $m_i \in \{0,\cdots,m\} \equiv M$ for $i = 1,\cdots,p$, and $\sum_{i=1}^{p} m_i = m$. The subvectors $\mathbf{s}_i$ of the preference vector s, where $i = 1,\cdots,p$ associate priorities i and goals s_{i,j_i}, where $j_i = 1,\cdots,m_i$, to the corresponding objective functions order j_i specified in the priority i. Note that the higher p indicates the higher priority. Example 8-1 below exemplifies the definition of the preference vector.

Example 8-1: Interpretation of preference vector

Consider the problem with three objectives, ψ_1,ψ_2,ψ_3. Assume the set of preference defined as follows:

- Priority level 1: $\psi_1 \geq 10, \psi_3 \geq 15$
- Priority level 2: $\psi_2 \geq 20$

Note that the priority level 2 is considered more important than the priority level 1. In this case, the preference vector can be specified as $\mathbf{s} = [(10,15),(20)]$ where $s_{1,1} = 10$ and $s_{1,1} = 15$ are associated with objective 1 and 3 with the lower priority compared to $s_{2,1} = 20$ which is associated with objective 2.

With the definition of the preference vector, the solution can also be defined in the same vector structure. Let $\mathbf{u} = \psi(\mathbf{x})$ be an m-dimensional objective vectors. $\mathbf{u}$ may be rewritten as:

$$\mathbf{u} = \left[\mathbf{u}_1, \cdots, \mathbf{u}_p\right] = \left[\left(u_{1,1}, \cdots, u_{1,n_1}\right), \cdots, \left(u_{p,1}, \cdots, u_{p,n_p}\right)\right].$$

With some permutation process, the subvector $\mathbf{u}_i$ can be defined such that the vector can be partitioned into two parts $\bar{\mathbf{u}}_i = \left\{u_{i,1}, \cdots, u_{i,k}\right\}$ and $\bar{\mathbf{u}}_i = \left\{u_{i,k+1}, \cdots, u_{i,m_i}\right\}$ where $u_{i,j} \in \bar{\mathbf{u}}_i \Leftrightarrow u_{i,j} \geq s_{i,j}$ and $u_{i,j} \in \bar{\mathbf{u}}_i \Leftrightarrow u_{i,j} < s_{i,j}$. The first part of $\mathbf{u}_i$, $\bar{\mathbf{u}}_i$, is the part that satisfies the goals set at the priority i whereas the second part, $\bar{\mathbf{u}}_i$, is the part not satisfying the corresponding goals. The same partition of an objective vector can also be applied to the other objective vector. For instance, let $\mathbf{v}_i$ be the other objective vector of level i for the objective vector v. $\bar{\mathbf{v}}_i^{\mathbf{u}}$ refers to the part of the vector $\mathbf{v}_i$ corresponding to the part of $\bar{\mathbf{u}}_i$. Example 8-2 illustrates the convention for the partition for the objective vector.

Example 8-2: Partition of objective vector

Consider the problem with four objectives, $\psi_1, \psi_2, \psi_3, \psi_4$. Assume the set of preference defined as follows:

- Priority level 1: $\psi_1 \geq 10, \psi_3 \geq 15, \psi_4 \geq 20$
- Priority level 2: $\psi_2 \geq 20$

From Definition 8-3, the preference vector can be defined as $\mathbf{s} = \left[(10,15,20),(20)\right]$. Let u be the objective vector with the values of 12, 20, 10, and 10 for $\psi_1, \psi_2, \psi_3, \psi_4$ respectively. Similarly, v is the other objective vector with the values of 15, 20, 20, and 5 for $\psi_1, \psi_2, \psi_3, \psi_4$ in that order. Thus, $\mathbf{u}_1 = (12,10,25)$ and $\mathbf{v}_1 = (15,20,5)$. From the definition of the vector partition given earlier, in order to partition $\mathbf{u}_1$, $\mathbf{u}_1$ needs to be reordered by switching the orders between ψ_3 and ψ_4. Then, $\mathbf{u}_1$ can be partitioned as follows: $\bar{\mathbf{u}}_1 = (12,25)$ and $\bar{\mathbf{u}}_1 = (10)$ where k

is 2. To apply the partition same operator from $\mathbf{u}_1$ to $\mathbf{v}_1$, $\mathbf{v}_1$ is reordered in the same way with $\mathbf{u}_1$, and now $\mathbf{v}_1 = (15,5,20)$. Then, $\bar{\mathbf{v}}_1^{\mathbf{u}} = (15,5)$ and $\bar{\mathbf{v}}_1^{\mathbf{u}} = (20)$.

With the definition of a more general definition of preference, the domination operator can be modified to $\underset{\mathbf{s}}{\succ}$ for defining a more general definition of non-dominated solution in Definition 8-1 where if $a \underset{\mathbf{s}}{\succ} b$, it means a dominates or is preferable to b given the preference vector $\mathbf{s}$. The definition of domination or preferability is defined next following Fonseca and Fleming (1998).

Definition 8-4: *Domination or Preferability*. An objective vector $\mathbf{u} = \left[\mathbf{u}_1, \cdots, \mathbf{u}_p\right]$ dominates or is preferable to $\mathbf{v} = \left[\mathbf{v}_1, \cdots, \mathbf{v}_p\right]$ under the preference vector $\mathbf{s} = \left[\mathbf{s}_1, \cdots, \mathbf{s}_p\right]$, denoted by $\mathbf{u} \underset{\mathbf{s}}{\succ} \mathbf{v}$, if and only if:

Case 1 with p = 1: $\left(\bar{\mathbf{u}}_p^{\mathbf{u}} >_p \bar{\mathbf{v}}_p^{\mathbf{u}}\right) \vee \left\{\left(\bar{\mathbf{u}}_p^{\mathbf{u}} = \bar{\mathbf{v}}_p^{\mathbf{u}}\right) \wedge \left[\left(\bar{\mathbf{v}}_p^{\mathbf{u}} < \bar{\mathbf{s}}_p^{\mathbf{u}}\right) \vee \left(\bar{\mathbf{u}}_p^{\mathbf{u}} >_p \bar{\mathbf{v}}_p^{\mathbf{u}}\right)\right]\right\}$

Case 2 with p > 1: $\left(\bar{\mathbf{u}}_p^{\mathbf{u}} >_p \bar{\mathbf{v}}_p^{\mathbf{u}}\right) \vee \left\{\left(\bar{\mathbf{u}}_p^{\mathbf{u}} = \bar{\mathbf{v}}_p^{\mathbf{u}}\right) \wedge \left[\left(\bar{\mathbf{v}}_p^{\mathbf{u}} < \bar{\mathbf{s}}_p^{\mathbf{u}}\right) \vee \left(\mathbf{u}_i \underset{\mathbf{s}_i}{\succ} \mathbf{v}_i \text{ for } i = 1, \ldots, p-1\right)\right]\right\}$

In the actual process, **u** and **v** are compared at the highest level of *p* specified by the user first. **u** dominates **v** if **u** meets more goals specified for priority level *p* than **v**, and there is no goal that **v** achieves and **u** does not at this priority level. If both of them achieve the same set of goals at level *p*, and also violate the other set of goals in exactly the same way $\left(\bar{\mathbf{u}}_p^{\mathbf{u}} = \bar{\mathbf{v}}_p^{\mathbf{u}}\right)$, then the next level of priority (*p* - 1) is considered. The process continues until it reaches the lowest priority level (*p* = 1) which will compare two vectors in a normal Pareto definition as explained in Definition 8-1. To clarify the concept, Example 8-3 is given below.

Example 8-3: Generalized definition of dominating solution

Consider again the problem stated in Example 8-2. Applying the generalized definition of the dominating solution given in Definition 8-4 above to compare **u** and **v** given the specified preference vector **s**:

Priority level 2: Both $\mathbf{u}_2$ and $\mathbf{v}_2$ satisfies the goal for this priority ($\psi_2 \geq 20$) and no other elements left to consider in this level so the process moves on to the next lower level.

Priority level 1: Now $\mathbf{u}_1=(12,10,25)$ and $\mathbf{v}_1=(15,20,5)$. $\mathbf{u}_1$ satisfies two goals of $\psi_1\geq 10$ and $\psi_4\geq 20$ where as v_1 satisfies two goals of $\psi_1\geq 10$ and $\psi_3\geq 15$. In this case, the condition stated for case 1 in Definition 8-4 is not satisfied. Thus, in this example **u** does not dominate v.

Two important lemmas can be deduced from the definition of the domination given (Fonseca and Fleming, 1998).

Lemma 8-1: For any two objective vectors **u** and **v**, if **u** $>_p$ **v**, then **u** either dominate or is equivalent to **v**, given any preference vector **s**.

Lemma 8-2 (*Transitivity*): The dominating relationship is transitive, i.e. given any three objective vectors, **u**, **v**, and **w**, and a preference vector **s**: $\mathbf{u}\underset{s}{\succ}\mathbf{v}\underset{s}{\succ}\mathbf{w}\Rightarrow\mathbf{u}\underset{s}{\succ}\mathbf{w}$.

These two lemmas illustrate the consistency and preservation of the key properties of the generalised definition of non-dominated solutions (Definition 8-4) compared to the traditional definition given in Definition 8-1. This implies that the conventional definition of Pareto solution is encompassed in Definition 8-4. If one defines a preference vector s with all objectives has equal priority (i.e. p = 1) and no goal levels are given (i.e. $\mathrm{s}=[\mathrm{s}_1]=[(-\infty,\ldots,-\infty)]$), then the definition of preferability or domination given in Definition 8-4 reduces to the conventional definition of Pareto preferability given in Definition 8-1.

The generalised definition of domination and preferability is not limited only to the case of the conventional definition of Pareto solution but can also be applied to other preference forms (see Section 8.2.1) by defining different forms of the preference vector s. For example, if s is defined with only one level and contains a goal for each objective, $\mathrm{s} = [\mathrm{s}_1,\ldots,\mathrm{s}_n] = [(\mathrm{s}_{1,1}),\ldots,(\mathrm{s}_{n,1})]$, then Definition 8-4 reduces to the definition of a preferable solution for the Goal programming.

The other possible utilisation of this generalised definition is for forming a constrained optimisation program. Let $g_l\geq s_l$ $l=1,\ldots,n$ be n inequality constraints and the main objective function of this problem is to maximise m objective functions. The constrained optimisation problem can be recast as the multiobjective problem with the preference vector of $\mathbf{s} = [\mathbf{s}_1, \mathbf{s}_2] = \left[(\infty_{1,1},\ldots,\infty_{1,m}),(s_{2,1},\ldots,s_{2,n})\right]$ where the constraints are converted into higher priority goals ($p = 2$). The same formulation can also be applied to a normal single objective constrained optimisation as discussed in the previous chapter.

With this generalised definition for the non-dominated solution (i.e. Pareto solution), the next section describes the extension of NSGA-II to the case of progressive preference articulation. The simple idea for forcing the NSGA-II to focus on a particular part of the Pareto front is introduced based on user's defined goals.

8.5.2 Progressive preference articulation for NSGA-II

As discussed in Section 8.2.2, the posterior preference articulation method (which is the original NSGA-II method proposed in Section 8.4) may waste the computational effort generating the trade-off surface outside the area of interest of the decision maker. An idea proposed in this section is to include the interaction process between the decision maker (transport planner) and the NSGA-II to enhance the focus the effort of NSGA-II on the area of most interest of the decision maker.

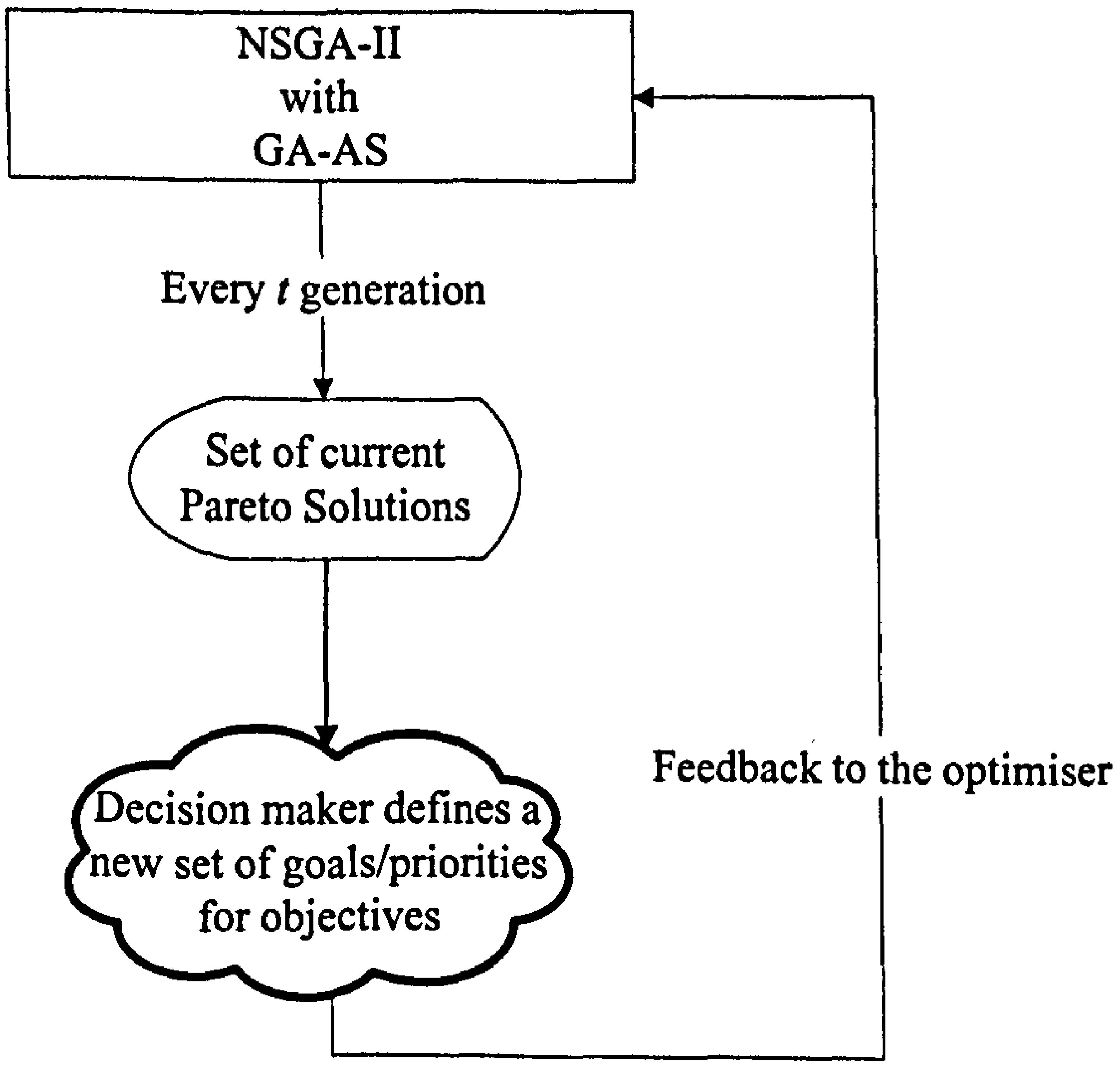


Figure 8-4 Framework for the progressive preference articulation method with NSGA-II

There are two types of preference type decision maker is allowed to supply to the NSGA-II given the current information from the NSGA-II to control the focus area of the search. These are goals and priorities. Figure 8-4 depicts roughly the overall idea of the interaction between the decision maker (DM) and the optimiser (NSGA-II with GA-AS). The decision maker will initially set the interruption interval in terms of generation number, i.e. decision maker wants to intervene with the optimiser every t generations of the evolution process. Then, after each t generations, the set of the current Pareto solutions will be displayed to the decision maker on-line and the decision maker will decide on which area of the current, possibly sub-optimal, Pareto front (not the true Pareto front yet) he/she wishes the optimiser to explore further. The input from the decision maker to the optimiser will be in a form of set goals and/or priorities. After receiving an additional preference vector from the decision maker, the optimiser proceeds the evolution process with additional goals and priorities.

Example 8-4 below demonstrates the intention of the interactive process explained above where the decision maker can supply the optimiser a set of additional goals and priorities during the process.

Example 8-4: Influence of progressive preference articulation on the GA search

Figure 8-5 below exemplifies the intent and process of the interactive goal setting for the preference articulation and focusing the search area. In this example, the problem is to find the solution maximising two objective functions (f_1 and f_2). Note that all the results shown in Figure 8-5 are purely for illustration purpose. Assume that the decision maker did not pre-define any goal for the first 50 generations of the NSGA-II. The decision maker wished to intervene the optimisation process every 50 generations of GA (t = 50). Thus, after 50 generations, the set of current non-dominated solutions as presented to the decision maker (figure on the top-left), the decision maker then decided to set the goal for f_1 and f_2, that are $s_{1,1}$ and $s_{1,2}$ shown on the left figure. Notice the subscript 2 for the priority for both $s_{1,1}$ and $s_{1,2}$. Note that the distribution of chromosomes (both dominated and non-dominated) are widely spread over the possible range of the pair of f_1 and f_2. The shaded cone in the left figure is the focuses region for the next 50 generations of the optimisation process as the results of the input goals. The focused region is the region of the Pareto front that GA will concentrate on.

After that, the optimisation algorithm continues to run for a further 50 generations. The second figure (top-right) shows the results at the 100th generation of the evolution process. What is expected to happen as the result of additional constraints from $s_{1,1}$ and $s_{1,2}$ is that the optimisation algorithm would spend more effort on the focused region (shaded cone). Thus, there should be more chromosomes appearing in this shaded cone. Similarly, the set of the non-dominated solutions should be totally inside the shaded cone (those chromosomes satisfy the goals set). At this stage, the decision maker is presented with the information on the current non-dominated solutions again. Then, the decision maker expresses any adjustment on goals. In the right figure, it is assumed that the decision maker sets the new goal as well as giving the priority to f_1. The evolution process will carry on as previously but concentrating more effort on this new focused region and on maximising f_1. The third figure (bottom-left) shows what could happen when an additional set of constrains (goals) and priorities is introduced. In this case, the optimiser is driven toward the point with extreme value of f_1 within the focused region.

The actual modifications needed for the NSGA-II presented in Section 8.4 are (i) the non-dominated sorting process (also referred to as the ranking process), (ii) the crowding tournament selection operator, and (iii) the intervention phase of the decision maker. The actual change for (i) and (ii) is relative simple where the operator will consider the new definition of the non-dominated solution (Definition 8-4) instead of the original definition of non-dominates solution given in Definition 8-1. The method of progressive preference articulation presented in this section will be tested with the network of Edinburgh later on in Section 8.6.3.

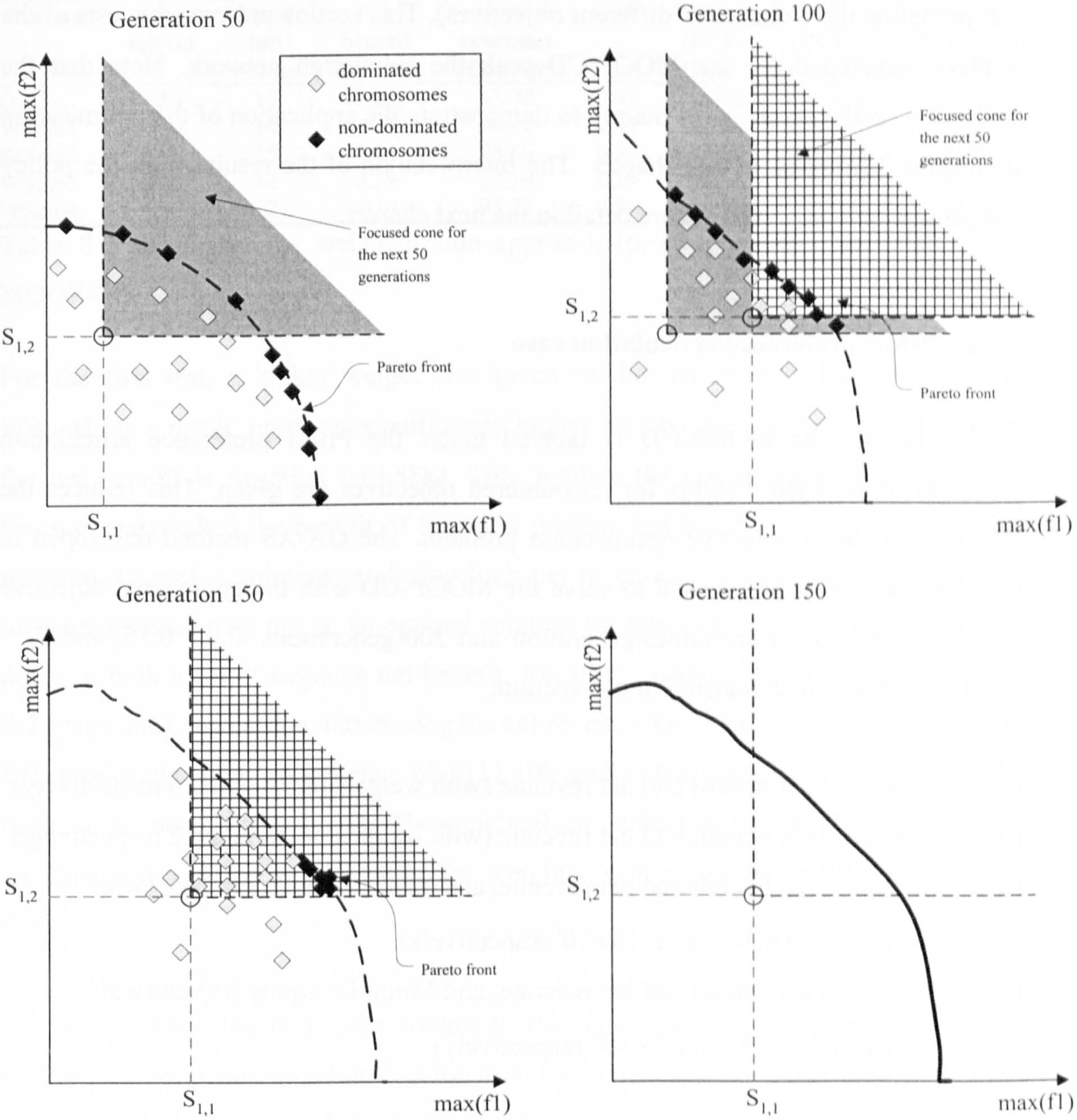


Figure 8-5 Illustration of focusing effect on the search as the results from decision maker's goals set

8.6 TESTS OF NSGA-II WITH EDINBURGH NETWORK

There are three possible frameworks for solving the MOOP, including Priori preference articulation, Posterior preference articulation, and Progressive preference articulation (See Section 8.2.2). The method of NSGA-II was developed both for the Posterior and Progressive preference articulation methods. In addition, the method developed in Chapter 6 can be adopted to solve the MOOP-CD with Priori preference articulation (i.e. predefine the weights for different objectives). This section presents the tests of the methods developed for the MOOP-CD with the Edinburgh network. Note that the discussion on the results aims mainly to demonstrate the application of the optimisation algorithms developed in this chapter. The interpretation of the results from the policy perspective will be given in more detail in the next chapter.

8.6.1 Priori preference articulation case

In this section the MOOP-CD is tackled under the Priori preference articulation paradigm. Predefined weights for encountered objectives are given. This reduces the MOOP to a single objective optimisation problem. The GA-AS method developed in Chapter 6 can then be adopted to solve the MOOP-CD with the aggregated objective function (with 50 chromosomes/generation and 200 generations, $P_m = 0.15$, and $P_c = 0.75$). Four tests are conducted in this section:

(i) Maximise net benefit and net revenue (with weights of 0.4 and 0.6 respectively)
(ii) Maximise net benefit and net revenue (with weights of 0.8 and 0.2 respectively)
(iii) Maximise net benefit and net revenue, and Minimise equity impact (with weights of $0.8, 0.2,$ and 10×10^6 respectively)
(iv) Maximise net benefit and net revenue, and Minimise equity impact (with weights of $0.5, 1,$ and 5×10^6 respectively)

Notice the difficulty in setting appropriate weights for the case with three objectives. The scale of the equity impact (*Gini* coefficient) is so different from the scale of the other two objectives (net benefit and net revenue). Hence, setting weights to reflect the intended trade-off can be complicated.

The GA-AS method is applied to all problem. Table 8-1 below shows the results from all tests with different weighted sum objective functions. All four optimal cordons found for the four tests (Wsum1 – Wsum4) are shown in Figure 8-6.

Cordon	Optimal toll (£)	No of toll points	Net benefit (£k/hour)	Net revenues (£k/hour)	*Gini*	TTtime (PCU-Hr/hr)	TTdis (PCU-Km/hr)	Demand (PCU/hour) '000
OPC1	1.50	13	7.21	43.70	0.41	52325	1720864	103.6
WSum1	4.00	14	-4.58	96.41	0.34	47218	1636791	97.94
WSum2	3.00	15	3.05	78.66	0.35	48827	1662902	99.70
WSum3	3.00	14	3.29	77.29	0.29	18891	1663228	99.71
WSum4	4.00	12	-3.61	95.23	0.34	47370	1633684	97.60

Table 8-1 Results of the weighted sum approach (priori preference articulation) for MOOP-CD

For the first test, a higher weight was given to the objective of net revenue. The WSum1, as a result, produces significantly higher net revenues (around £96k) whereas the net benefit is negative (-£4.58k). This implies the contribution of the high net revenue outweighed the benefit of having a positive net benefit. Hence, the algorithm just tried to seek a solution producing high net revenue. However, a cordon with too high net revenue may not be an optimal solution for this weighted sum problem either, due to a high level of negative net benefit. Recall the results from Chapter 6 in Table 6-2 (page 206), the cordon maximising the net revenue, OPC-REV (with the toll level of £4), produced a high net revenue of £111.10k and at the same time also generated a negative net benefit of £-49.19k. The weighted sum objective for this problem (with the weights of Wsum1) is around £48k. The weighted sum objective for WSum1 is around £56k.

The second test allocated more weight to the objective of net benefit. The optimal uniform toll level for the WSum2 is found to be £3, just slightly lower than the optimal uniform toll for WSum1. With the revised weights, the solution for WSum2 has a positive net benefit of around £3k and the net revenue of £79k. The net benefit is still substantially lower than the net benefit from the OPC1 (around £7k). This implies the cordon design with a relatively high net revenue still performs well against the new weights adopted.

The third and fourth tests are also similar in pattern to the previous two tests. The fourth test allocated more weight to the objective of net revenue compared to the third test. Hence, the WSum4 generated a higher net revenue (£95k) compared to the WSum3 (£77k). The optimal toll level for the WSum4 is also higher than that of WSum3 (similar to the comparison between WSum1 and WSum2).

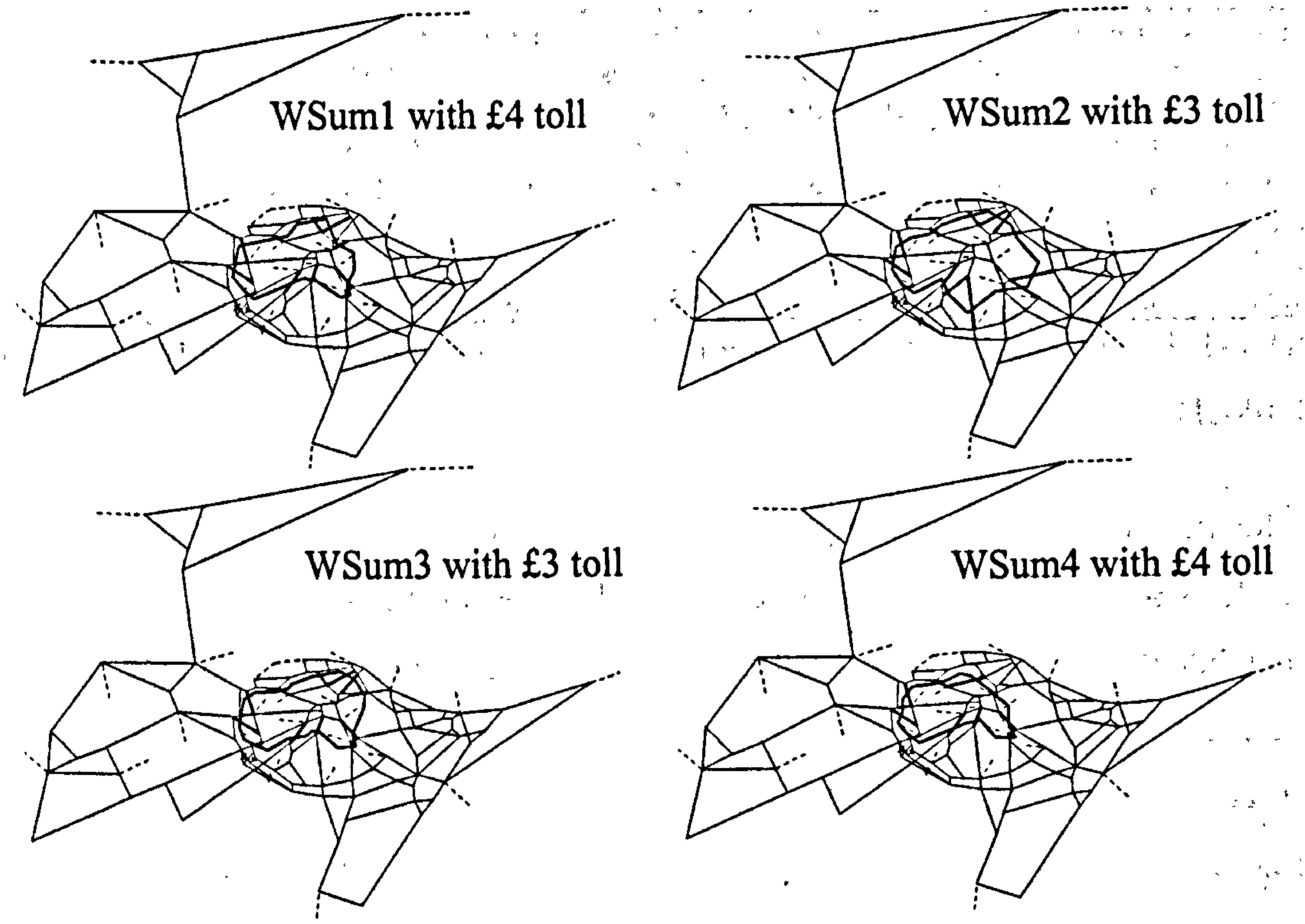


Figure 8-6 Results for MOOP-CD with weighted sum approach (highlighted links are toll links)

The results from all four tests show the domination of the objective of maximising net revenue. The reason is due to the scale of the objective (i.e. the level of revenue is always substantially higher than the net benefit). The results also demonstrated the point made earlier about the potential difficulty with the weighted sum approach (or in general the priori preference articulation approach), since the decision maker cannot know the actual trade-off between the different objectives prior to the process of determining the weights or preferences. Also, as demonstrated in the results, the distribution of weights does not necessarily coincide with the distribution of the non-dominated solutions. This problem can be observed further with the results in Section 8.6.2.

In terms of the behaviour of the optimisation algorithm, Figure 8-7 and Figure 8-8 show the elements of the best solution found so far during the GA-AS process. The figures show the net benefits, net revenues, and the weighted sum objective for the best solution found up to stage of the optimisation process (different generations of the GA-AS).

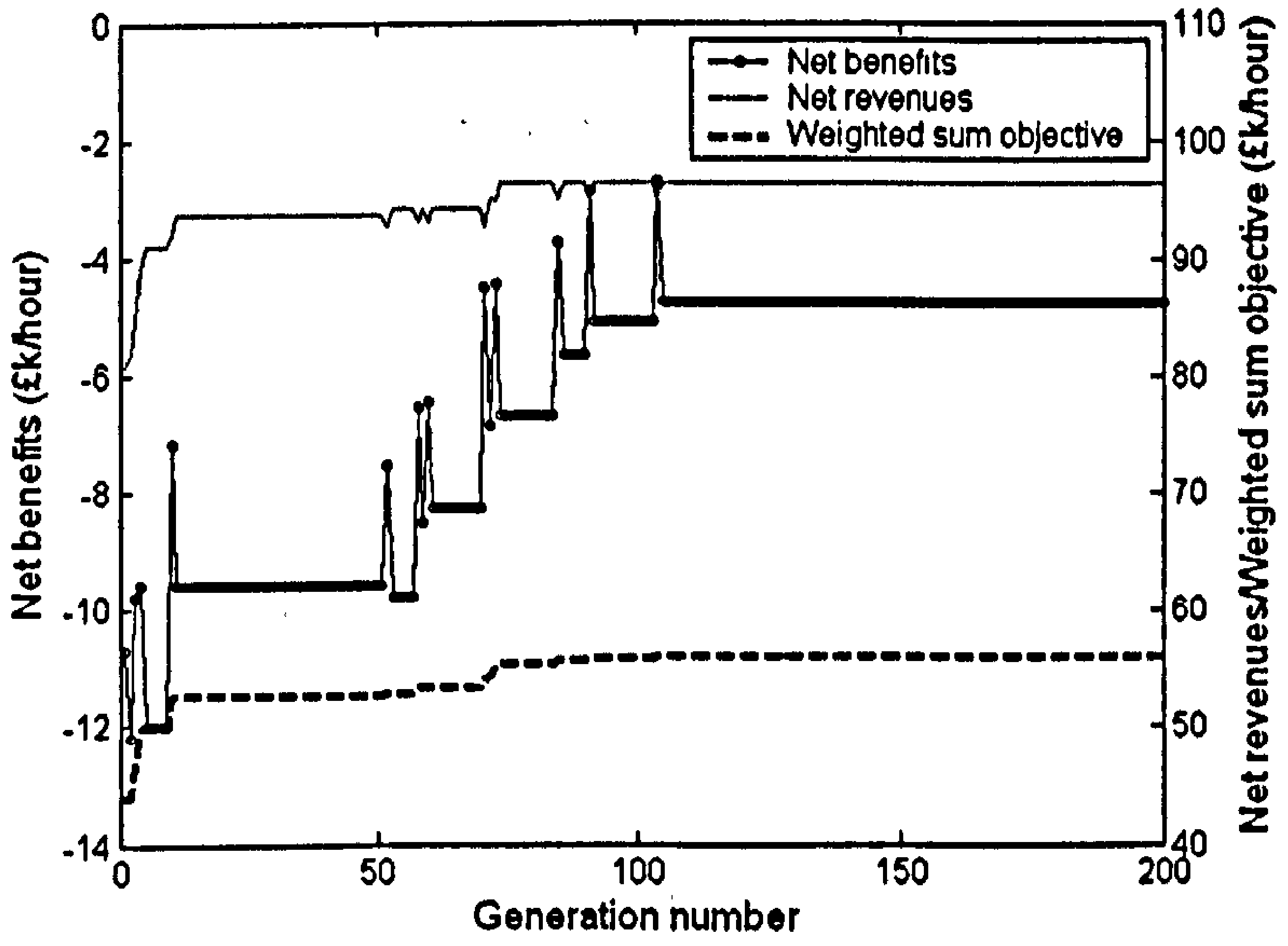


Figure 8-7 Elements of the best solution found so far for the weighted sum problem (WSum1)

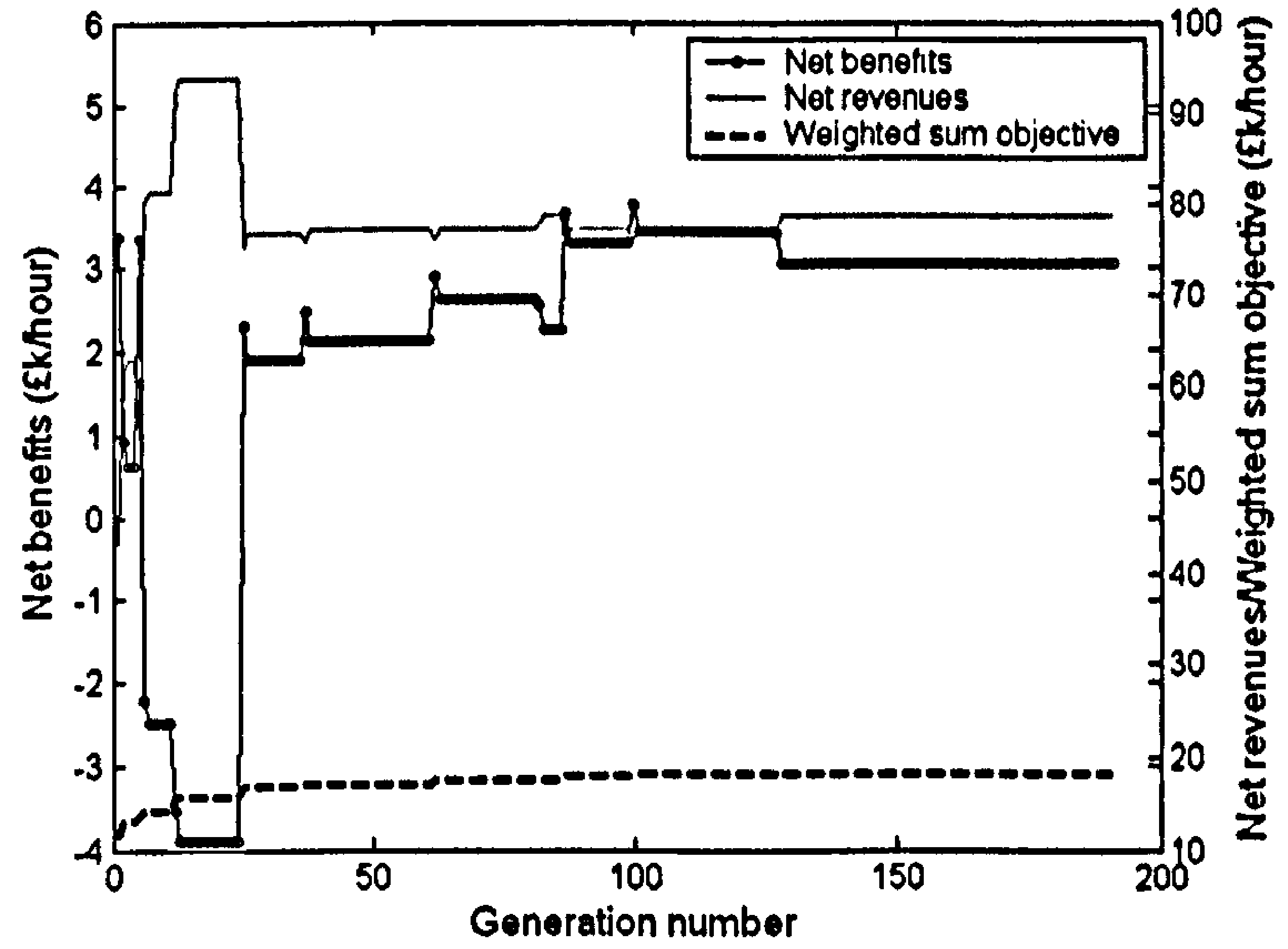


Figure 8-8 Elements of the best solution found so far for the weighted sum problem (WSum2)

Both figures demonstrate the attempt of the algorithm to find the right balance between both objectives (net benefit and net revenue). During the optimisation process, there exist several blips in the level of net benefit and net revenue leading toward the stabilised levels in later generations. The result in Figure 8-7 shows that the algorithm can pick up straightaway that the optimal solution for this problem must consist of a high level of net revenue (due to its high weight). The algorithm then gradually tried to increase both the net benefit and net revenue. The blips on the net benefit curve showed the situation when the algorithm tried a solution with a bit higher level of the net benefit. Then, the algorithm recognised the greater advantage of increase the net revenue rather than the net benefit in relation to the improvement over the weighted sum objective function.

On the contrary, the result in Figure 8-8 tells us a slightly different story. Test WSum2 gave more weight to the objective of maximising the net benefit. The net benefit for the best solution found so far in Figure 8-8 increased instantly beyond £3k in early generations whereas the net revenue is significantly lower than the net revenue during the same stage for the WSum1. Then, the algorithm started to seek the solution with higher net revenue. At the period when the net revenue reached it highest value (from the overall process), the net benefit dropped to almost £-4k (around 25th generation). However, at this point the algorithm found a better compromise between the two objectives by raising the net benefit and sacrificing some of the net revenue. This is due to the pre-defined weights. Then, a similar phenomenon occurring with the WSum1 test also came out during the optimisation process with the WSum2 test where the algorithm tried to increase the net benefit to seek a better compromise. Finally, the algorithm converged to a stabilised compromise between the two objectives.

8.6.2 Posterior preference articulation case

In this section, the Posterior preference articulation approach is applied to MOOP-CD. The method adopted for generating the non-dominated solutions is the NSGA-II (with GA-AS) as explained in Section 8.4. The difficulty in setting appropriate weights for the Priori preference articulation approach was mentioned in the previous section. As explained in Section 8.2, the benefit of Posterior preference articulation based approach is that the decision maker is presented with the actual information about the real trade-off between different objectives (from the set of non-dominated solutions). The problem of setting appropriate weights is thus eliminated with the Posterior preference articulation based approach. NSGA-II developed in 8.4 is used in conjunction with the GA-AS method to solve the MOOP-CD (with 50 chromosomes/generation and 200 generations, $P_m = 0.15$, and $P_c = 0.75$). The tests presented in this section are as follows:

(i) Maximise net benefit and net revenues
(ii) Maximise net benefit and Minimise equity impact

(iii) Maximise net benefit and Minimise total travel time

(iv) Maximise net benefit and Minimise total travel distance

(v) Maximise net benefit and net revenue, and Minimise equity impact

(i) Maximising net benefit and net revenues

The solution being sought in MOOP is a set of Pareto solutions creating a Pareto front. After applying NSGA-II to the problem, Figure 8-9 shows the solutions produced and Pareto front found by the algorithm for the first test (maximise net benefit and net revenue). The grey dots represent different solutions generated during the process of NSGA-II. The black dots are the Pareto solutions which are used to create the Pareto front (black line).

The NSGA-II method generated totally 31 non-dominated solutions whose the levels of net benefit and net revenue (per peak hour) ranged from £-3.6k to £7.2k and from £43.7k to £95.2k respectively. Table 8-2 contains the performance indicators for all 31 Pareto solutions. The first solution (WR1) is the solution with the highest net revenue (top solution of the Pareto front). The other solutions are named sequentially by their order on the Pareto front. Based on this particular test, there exists a conflict between the two objectives considered; in other words, the gain in either net revenue or net social welfare improvement must be inevitably traded off by the loss in the other objective. There are some solutions capable of producing the positive net benefit and net revenue (solutions WR12 -WR31). The higher the uniform toll level the higher the ability of a cordon to generate the revenue. On the other hand, the cordon with high toll may not yield a good result in terms of the net benefit.

In theory, the two solutions found in Section 8.6.1 should also be found as one of the non-dominated solutions. However, these two solutions were not actually found during the process of NSGA-II. This is not a surprise given the complexity of the problem and the computational effort given. The WSum1 solution is just over the WR1 solution with a higher net revenue and lower net benefit. The WSum2 solution is in between WR13 and WR14.

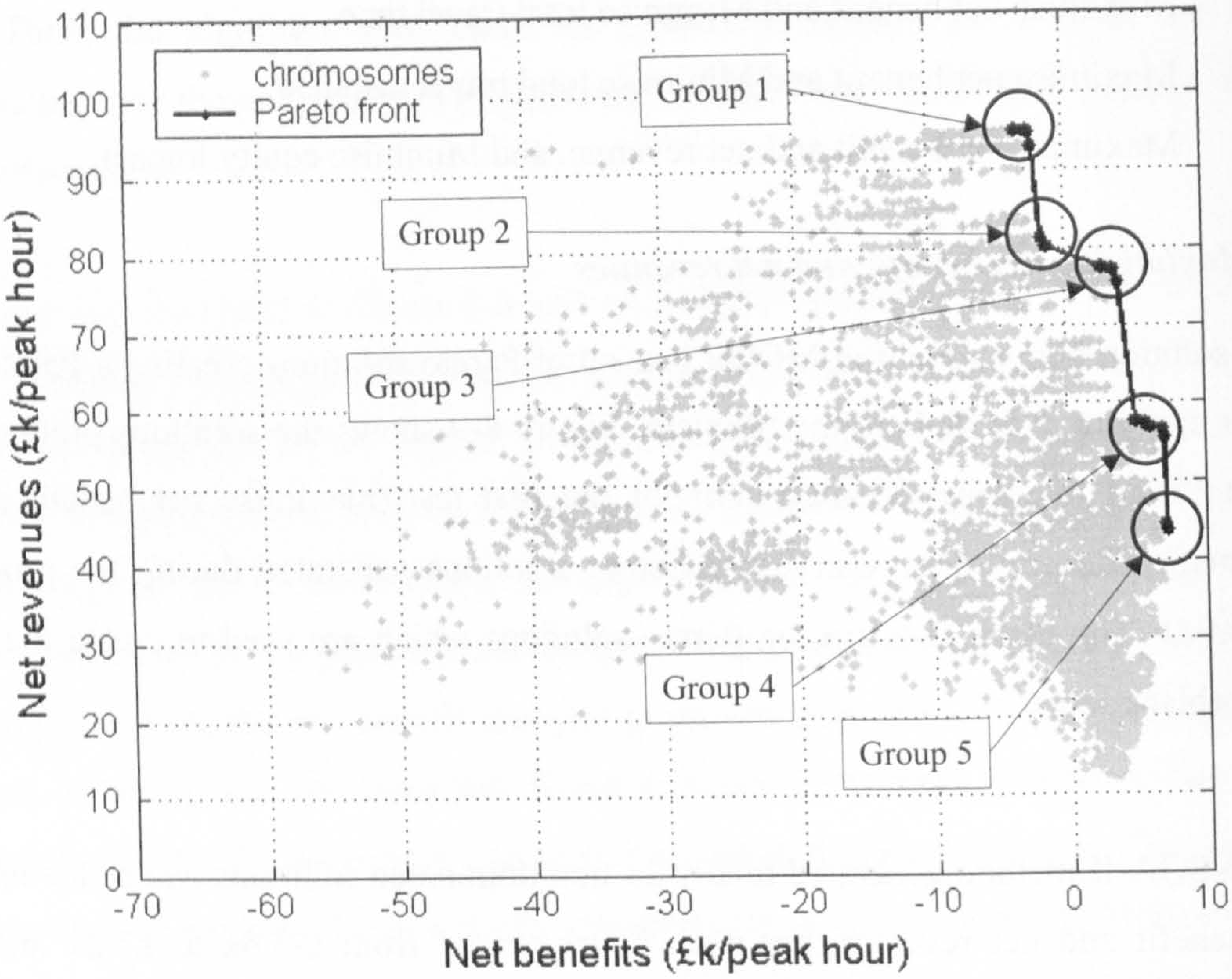


Figure 8-9 Pareto solutions and corresponding Pareto front for MOOP problem with objectives of maximising net benefits and net revenues

When the planner is presented with these possible solutions and their performances/design (information from Table 8-2), he or she can then make a selection on the best compromised design. This is the underlying concept of posterior preference articulation approach.

As depicted in Figure 8-9, 31 non-dominated solutions can be grouped into five different clusters based upon their objectives. If the Priori preference articulation paradigm is applied to this decision making (See Section 8.6.1), the weighted sum objectives (for different weight pairs) can be produced based on the results in Table 8-2. Figure 8-10 shows the weighted sum objective from the results in Table 8-2 with different weight pairs. Six different weight pairs are applied to the solutions (the first weight in the legend of the figure is the weight for the net benefit, and the second is for the net revenue).

Solution No.	Toll (£)	No of toll points	Net benefit (£k/hr)	Net revenues (£k/hr)	TTtime (PCU-hr/hr)	TTdis (PCU-km/hr)	Total demand	*Gini*
WR1	4.00	12	-3.61	95.24	47370.00	1633684.00	97603.00	0.34
WR2	4.00	12	-3.47	95.23	47344.00	1634462.00	97644.00	0.33
WR3	4.00	14	-2.66	95.02	47218.00	1636791.00	97943.00	0.34
WR4	4.00	14	-2.60	95.00	47201.00	1636937.00	97965.00	0.35
WR5	4.00	14	-2.57	94.99	47194.00	1637303.00	97968.00	0.35
WR6	4.00	13	-2.30	93.04	47314.00	1649430.00	98277.00	0.39
WR7	4.00	13	-1.75	81.24	48581.00	1673092.00	100126.00	0.45
WR8	4.00	12	-1.42	80.05	48670.00	1680815.00	100256.00	0.51
WR9	4.00	12	-1.40	80.05	48666.00	1680716.00	100255.00	0.50
WR10	4.00	12	-1.35	80.04	48650.00	1680906.00	100274.00	0.51
WR11	4.00	12	-1.30	80.03	48640.00	1680896.00	100273.00	0.52
WR12	3.00	12	2.54	77.52	48996.00	1659782.00	99364.00	0.33
WR13	3.00	12	2.67	77.50	48959.00	1659905.00	99379.00	0.34
WR14	3.00	14	3.30	77.29	48891.00	1663228.00	99707.00	0.30
WR15	3.00	14	3.38	77.28	48862.00	1662646.00	99686.00	0.34
WR16	3.00	14	3.47	77.27	48849.00	1662903.00	99722.00	0.35
WR17	3.00	14	3.54	77.26	48824.00	1663042.00	99708.00	0.34
WR18	3.00	13	3.74	75.74	48919.00	1673876.00	99999.00	0.36
WR19	3.00	13	3.75	75.73	48910.00	1673434.00	99987.00	0.39
WR20	2.00	13	4.91	57.58	51051.00	1692619.00	100958.00	0.36
WR21	2.00	13	4.99	57.57	51030.00	1692681.00	100976.00	0.36
WR22	2.00	15	5.68	57.33	50933.00	1695643.00	101267.00	0.33
WR23	2.00	16	5.69	57.23	50902.00	1695591.00	101255.00	0.36
WR24	2.00	12	5.92	56.75	51109.00	1691651.00	101638.00	0.34
WR25	2.00	14	6.79	56.50	50963.00	1694784.00	101945.00	0.34
WR26	2.00	14	6.86	56.49	50950.00	1694995.00	101964.00	0.36
WR27	2.00	13	7.00	55.44	51018.00	1703384.00	102213.00	0.40
WR28	1.50	14	7.02	44.55	52287.00	1713528.00	103357.00	0.34
WR29	1.50	14	7.04	44.56	52283.00	1713466.00	103353.00	0.34
WR30	1.50	13	7.20	43.75	52326.00	1720870.00	103597.00	0.41
WR31	1.50	13	7.21	43.76	52325.00	1720864.00	103597.00	0.41

Table 8-2 Non-dominated solutions from the MOOP, maximising net benefit and net revenue

Five different clusters of the non-dominated solutions become the best compromised solutions with different weights. The first cluster of the solutions comes out as the best compromise solutions with weight pairs of (0.4, 0.6), (0.5, 0.5), (0.6, 0.4), and (0.7, 0.3). This cluster of solutions covers a wide range of the weights. The third cluster becomes the best compromise solutions when the weight pair of (0.8, 0.2) is applied. Finally, the fourth cluster becomes the optimal solution when the weight pair of (0.9, 0.1) is adopted. Of course, the last cluster will definitely become the best solution when an extreme weight of (1.0, 0.0) is used. This result explains the difficulty found with the weighted sum approach (see Section 8.6.1) where the objective of net revenue dominates a large span of the weights due to its scale compared to that of the net benefit. Again, as discussed in the review in Section 8.3 using a good spread of the weights to create the single objective optimisation problem and solve it does not guarantee the well distributed non-dominated solution. This demonstrates the benefit of the NSGA-II approach.

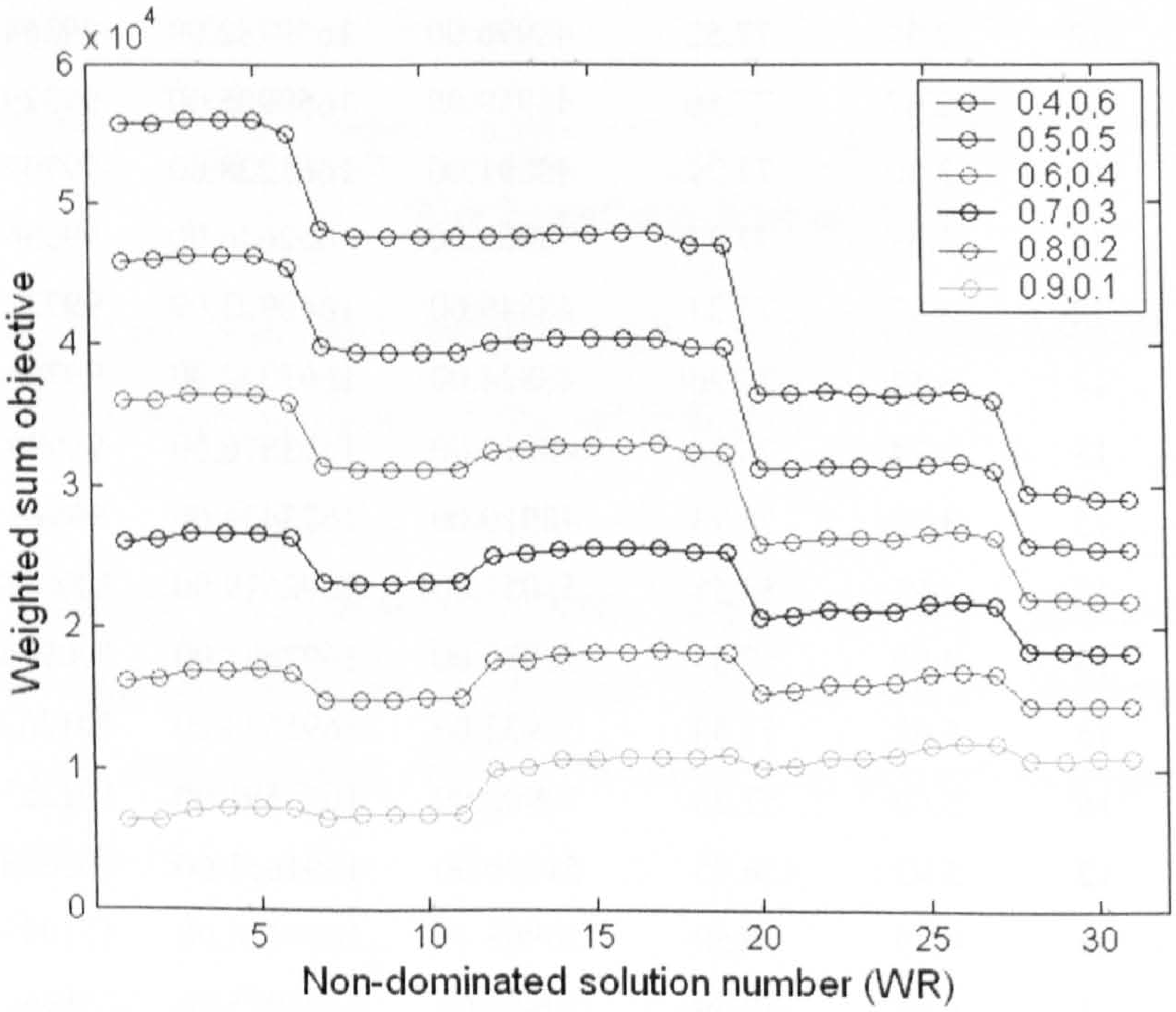


Figure 8-10 Weigthed sum objectives with different weight pairs (weight for net benefit, weight for net revenue)

(ii) Maximising net benefit and Minimising equity impact

For the second test, instead of maximising the net revenues, an objective function of minimising the *Gini* coefficient is adopted. Figure 8-11 shows the chromosomes generated by the NSGA-II process and the Pareto front identified. 31 Pareto solutions were found. They are presented in Table 8-3. The non-dominated solution with the lowest Gini coefficient (located at the bottom of the Pareto front) is named WG1, and the other solutions are named sequentially following the order of the Gini coefficients. The lowest level of *Gini* coefficient is 0.108 for the WG1 cordon but the net benefit drops to around £2.33k per peak hour compared to the net benefit from WG31 of £7.21k per peak hour.

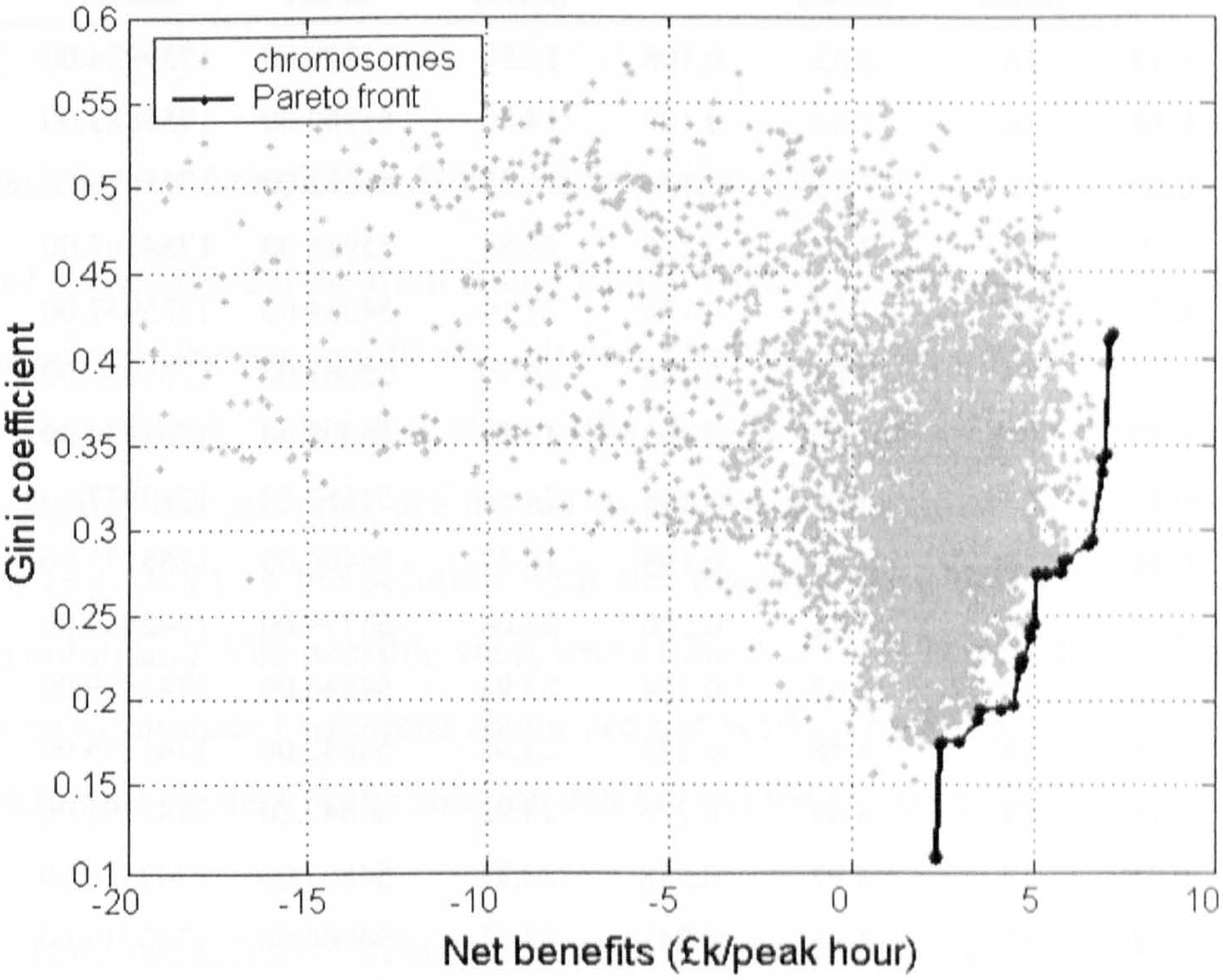


Figure 8-11 Pareto solutions and corresponding Pareto front for MOOP problem with objectives of maximising net benefits and minimising *Gini* coefficient (equity impact)

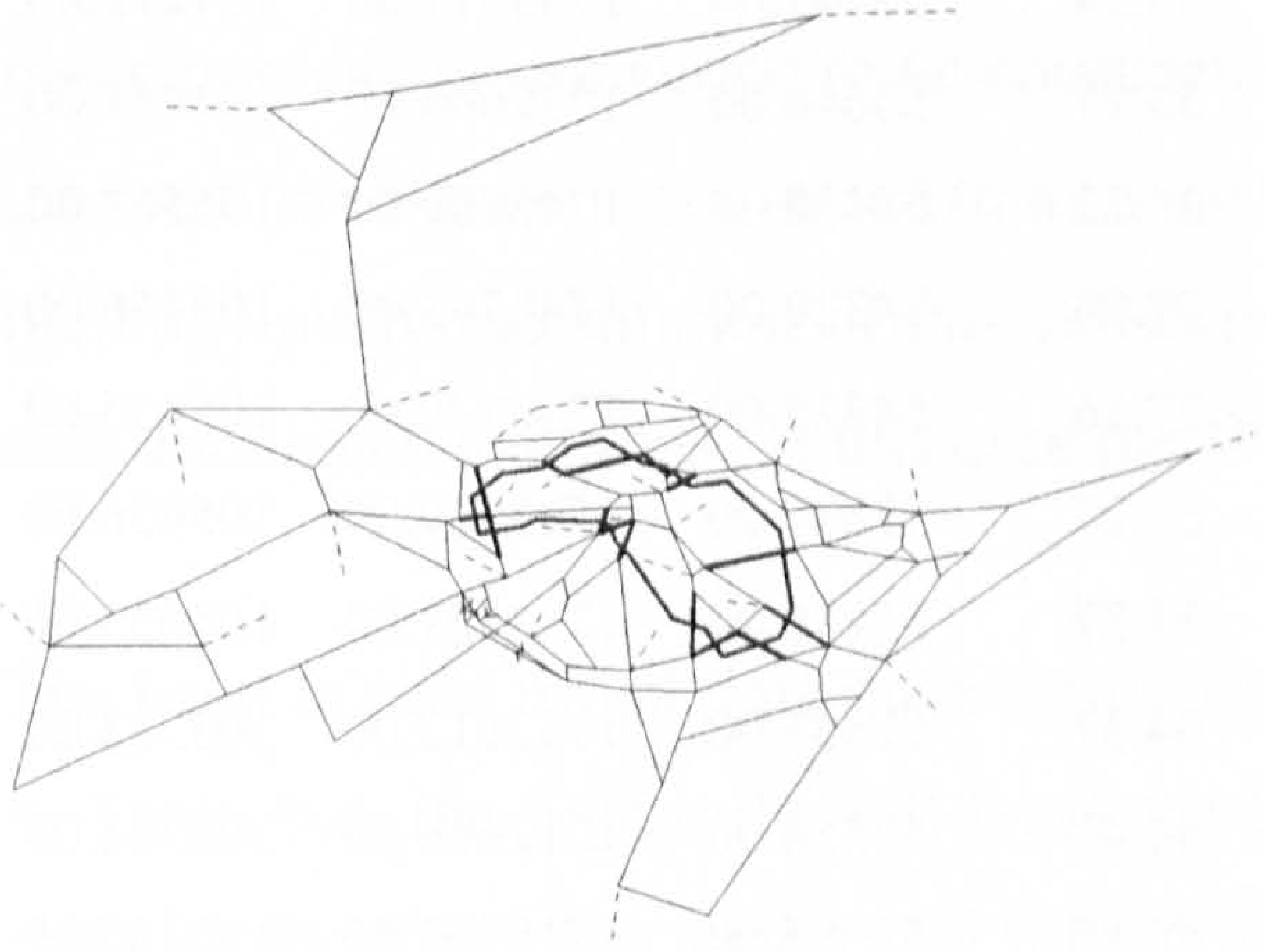


Figure 8-12 Location of WG1 cordon (with toll level of £0.50)

Figure 8-13 Location of WG31 cordon (with toll level of £1.50)

Figure 8-12 and Figure 8-13 show the locations of WG1 and WG31 respectively. WG31, which is the cordon with the highest net benefit, is unsurprisingly the same cordon as OPC1. The toll level for WG1 is relatively low (£0.50) which is the common characteristic of those cordons in Table 8-3 with a low Gini coefficient. The net benefit of WG31 is almost four times the net benefit of WG1.

Solution No.	**Toll (£)**	**No of toll points**	**Net benefit (£k/hr)**	***Gini***	**Net revenues (£k/hr)**	**TTtime (PCU-hr/hr)**	**TTdis (PCU-km/hr)**	**Total demand**
WG1	0.50	16	2.33	0.108	14.80	55970.00	1759924.00	107144.00
WG2	0.50	16	2.33	0.109	14.80	55969.00	1759883.00	107144.00
WG3	0.50	12	2.43	0.175	14.08	56072.00	1765470.00	107407.00
WG4	0.50	15	2.47	0.176	14.88	55956.00	1764167.00	107067.00
WG5	0.75	18	2.97	0.176	27.19	54584.00	1735957.00	105001.00
WG6	0.50	15	3.48	0.189	16.70	55661.00	1759085.00	106884.00
WG7	0.50	9	3.50	0.195	13.81	56098.00	1765834.00	107759.00
WG8	0.75	15	4.10	0.196	24.34	54857.00	1739977.00	105741.00
WG9	0.50	15	4.43	0.198	18.47	55406.00	1755173.00	106807.00
WG10	0.75	16	4.64	0.220	24.08	54779.00	1742889.00	105978.00
WG11	0.75	14	4.67	0.224	23.93	54846.00	1741829.00	105884.00
WG12	0.75	14	4.68	0.225	23.92	54842.00	1741796.00	105881.00
WG13	0.75	14	4.68	0.226	23.92	54842.00	1741791.00	105881.00
WG14	0.75	12	4.92	0.238	23.91	54863.00	1743558.00	105957.00
WG15	0.75	12	4.93	0.243	23.91	54860.00	1743519.00	105955.00
WG16	0.75	11	5.04	0.245	24.01	54857.00	1743518.00	105955.00
WG17	1.00	15	5.05	0.274	34.24	53546.00	1734395.00	104272.00
WG18	1.00	14	5.15	0.274	34.34	53547.00	1734411.00	104272.00
WG19	1.00	14	5.28	0.274	33.17	53646.00	1737066.00	104451.00
WG20	0.75	14	5.38	0.275	26.22	54430.00	1745809.00	105562.00
WG21	0.75	15	5.69	0.276	27.06	54329.00	1740763.00	105456.00
WG22	0.75	14	5.79	0.276	27.16	54328.00	1740770.00	105455.00
WG23	1.00	15	5.84	0.282	31.27	53887.00	1733839.00	105056.00
WG24	1.00	14	5.94	0.283	31.37	53886.00	1733821.00	105055.00
WG25	1.50	16	6.53	0.292	44.37	52351.00	1712612.00	103333.00
WG26	1.50	15	6.62	0.294	44.47	52354.00	1712666.00	103334.00
WG27	1.50	15	6.93	0.334	44.45	52286.00	1713497.00	103355.00
WG28	1.50	15	6.94	0.343	44.45	52283.00	1713459.00	103353.00
WG29	1.50	14	7.03	0.345	44.55	52290.00	1713807.00	103379.00
WG30	1.50	14	7.10	0.411	43.65	52327.00	1720869.00	103598.00
WG31	1.50	13	7.21	0.414	43.75	52325.00	1720864.00	103597.00

Table 8-3 Non-dominated solutions from the MOOP, maximising net benefit and minimising gini coefficient

From Table 8-3, some interesting relationship between the cordon design and its performance against different objective functions can be observed. There seems to be a strong connection between the performances of the cordons against the equity objective and the level of their uniform tolls. From the results, it is likely that the lower the toll level the more equitable the cordon scheme. This issue will be clarified further in Chapter 9.

(iii) Maximising net benefit and Minimising total travel time

Figure 8-14 shows the Pareto front found by the NSGA-II for the third tests. The third test is to maximise the net benefit and minimise the total travel time. 14 non-dominated solutions were found. Table 8-4 shows the details for all of these non-dominated solutions. Again, the solutions are named in the ascending order of the level of total travel time (e.g. WT1 is the solution with the lowest total travel time from all non-dominated solutions). The possible weakness of the NSGA-II is revealed in Figure 8-14 where the non-dominated solutions found are not well spread over the Pareto front. In fact, five different clusters of the solution can be defined (noted in the figure).

The result from the CON-TTTIME discussed in Section 7.4.2 in the previous chapter also appeared to be one of the non-dominated solutions (WT9). This followed one of the classical multiobjective optimisation approaches as explained in Section 8.3.1 where the objectives can be converted to be constraints. In this case, if the objective of minimising the total travel time is converted to a constraint as depicted as a dash line in Figure 8-14, then obviously the solution of this constrained optimisation problem will be one of the non-dominated solutions (in this case the WT9).

The level of total travel time does seem to be related to the net revenue. From the results in Table 8-4, the higher the toll level the higher the net revenue (as discussed earlier) and also the lower the total travel time. This can probably be explained by the fact that in the most extreme case if all links are tolled with a very high toll, eventually there will be no traffic left in the network. Hence, the total travel time reaches its minimum possible level, zero.

Solution No.	Toll (£)	No of toll points	Net benefit (£k/hr)	TTtime (PCU-hr/hr)	Net revenues (£k/hr)	*Gini*	TTdis (PCU-km/hr)	Total demand
WT1	4.00	14	-2.57	47194	94.996	0.349	1637303	97968
WT2	4.00	14	-2.39	47311	92.939	0.390	1649449	98276
WT3	4.00	13	-2.30	47314	93.039	0.389	1649430	98277
WT4	4.00	15	-1.93	48570	81.054	0.458	1672697	100111
WT5	4.00	14	-1.83	48571	81.154	0.457	1672706	100111
WT6	4.00	13	-1.75	48581	81.244	0.445	1673092	100126
WT7	4.00	12	-1.30	48640	80.038	0.520	1680896	100273
WT8	3.00	14	3.54	48824	77.267	0.345	1663042	99708
WT9	3.00	13	3.75	48910	75.735	0.394	1673434	99987
WT10	2.00	15	6.77	50948	56.397	0.364	1694971	101962
WT11	2.00	14	6.86	50950	56.497	0.361	1694995	101964
WT12	2.00	13	7.00	51018	55.441	0.398	1703384	102213
WT13	1.50	14	7.04	52283	44.551	0.340	1713466	103353
WT14	1.50	13	7.21	52325	43.752	0.414	1720864	103597

Table 8-4 Non-dominated solutions from the MOOP, maximising net benefit and minimising total travel time

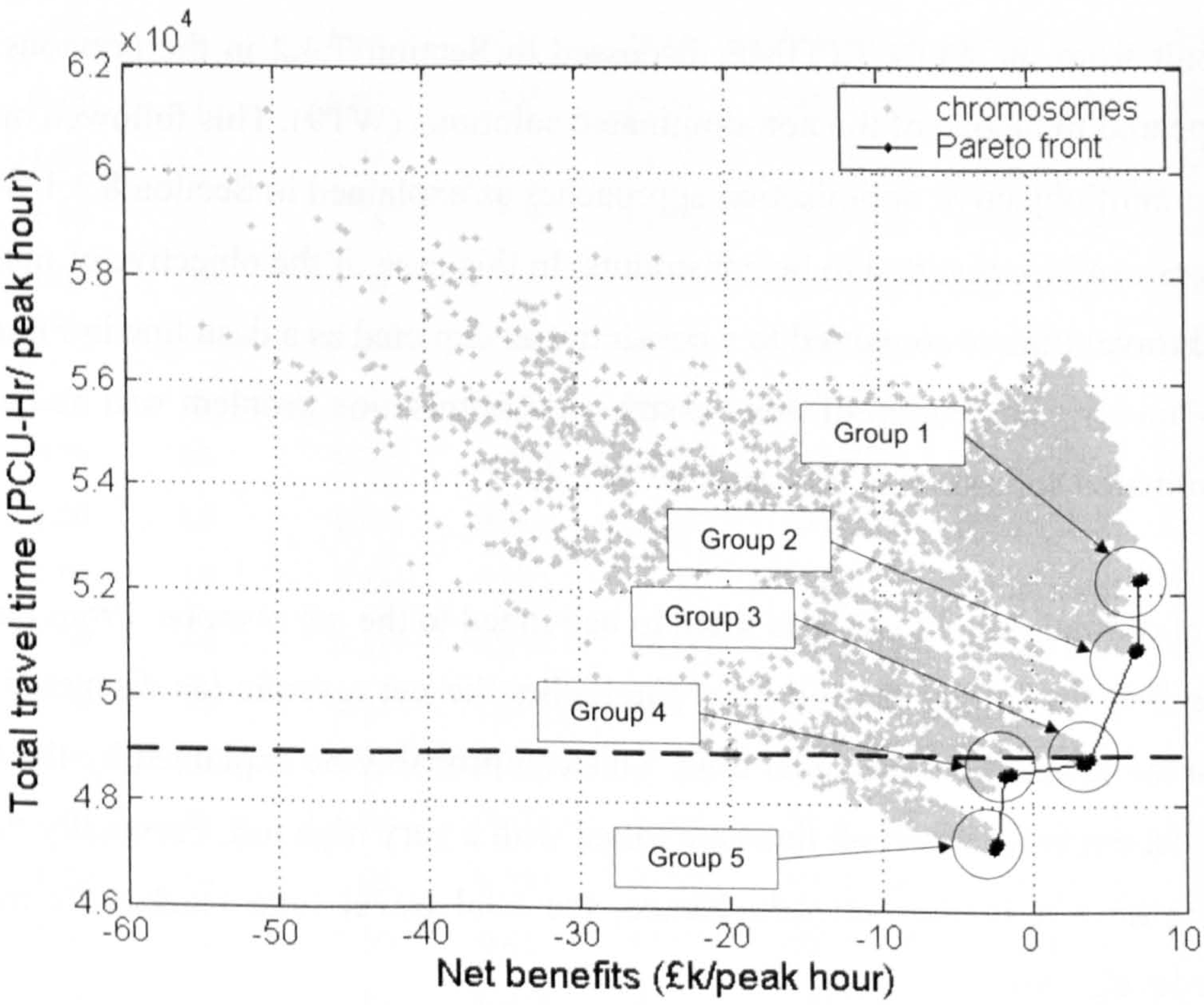


Figure 8-14 Pareto solutions and corresponding Pareto front for MOOP problem with objectives of maximising net benefits and minimising total travel time

Figure 8-15 shows the location of the WT1 cordon (the number of toll points is 14). Obviously, the WT1 cordon cover a wider range of the network compared to the OPC1 cordon. This means a higher number O-D pair or possible route may be included inside the charged zone. Thus, the cordon can impose the toll on a higher proportion of trips. With a high toll, the number of trips depressed and the level of congestion relieved can be achieved greatly. A wide coverage also kept the increase of travel time due to traffic diversion to a longer route (in terms of travel time) at the minimum. However, a too wide coverage area of a cordon may not achieve a good result in terms of reducing the total travel time if most of the trips actually happen inside the cordon and hence they are not affected by the toll at all.

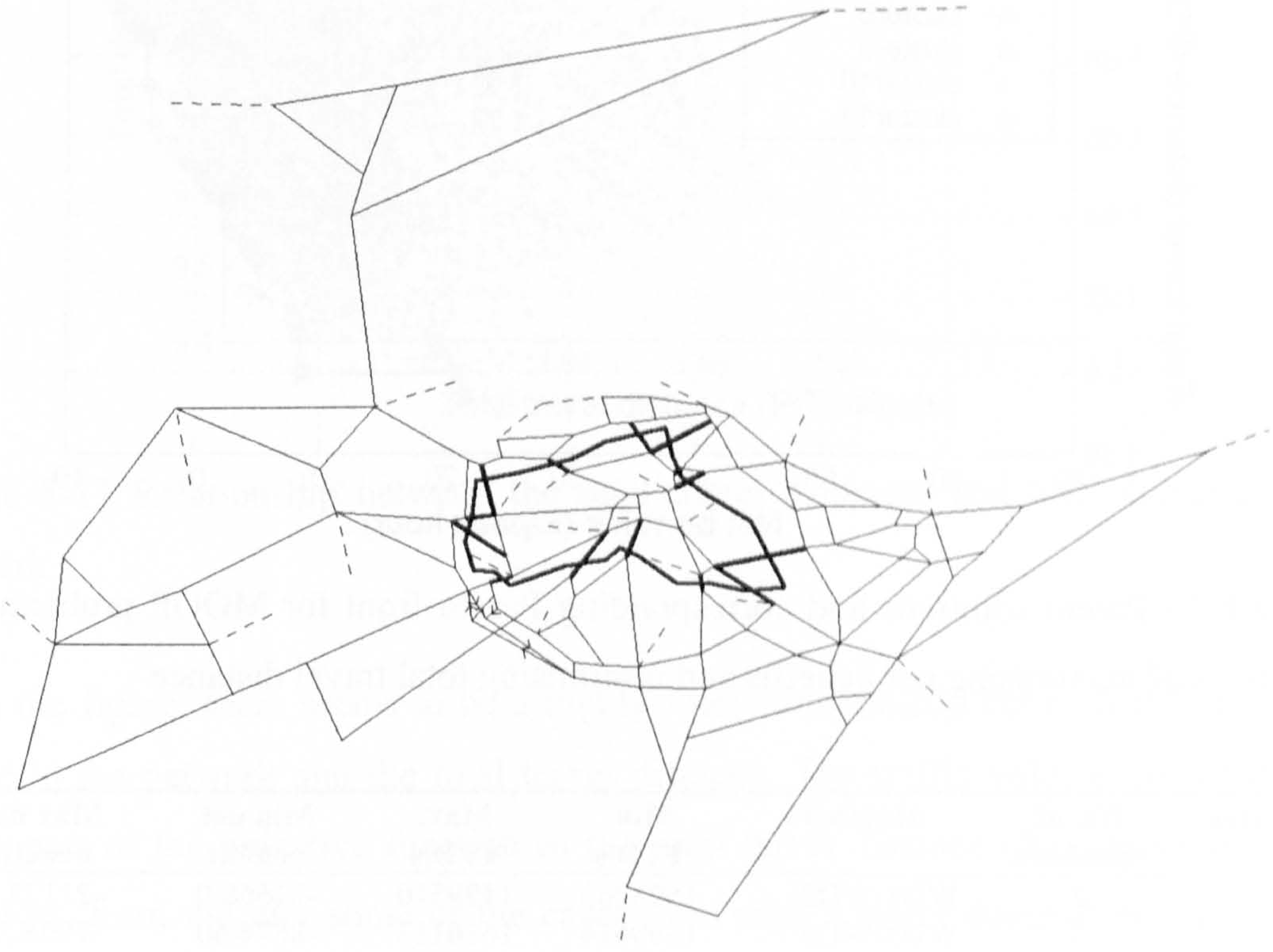


Figure 8-15 WT1 Cordon with the optimal toll level of £4

(iv) Maximising net benefit and Minimising total travel distance

Figure 8-16 shows the Pareto front found by the NSGA-II for the fourth test which aims to maximise the net benefit and minimise the total travel distance. 95 non-dominated solutions where identified. It is considered impractical to present the information for all 95 non-dominated solutions. Thus, the non-solutions are clustered into 11 different groups by their objectives (See Figure 8-16). Again, the non-dominated solutions are named following the ascending order of the total travel distance (i.e. WD1 is the solution with the lowest total travel distance and WD95 is the solution with the highest total travel distance). Table 8-5 shows the information of each cluster (ranges of the net benefit and total travel distance and the number of members in each cluster). Unlike the test discussed in the previous section, the algorithm found a very nice well spread set of the non-dominated solutions.

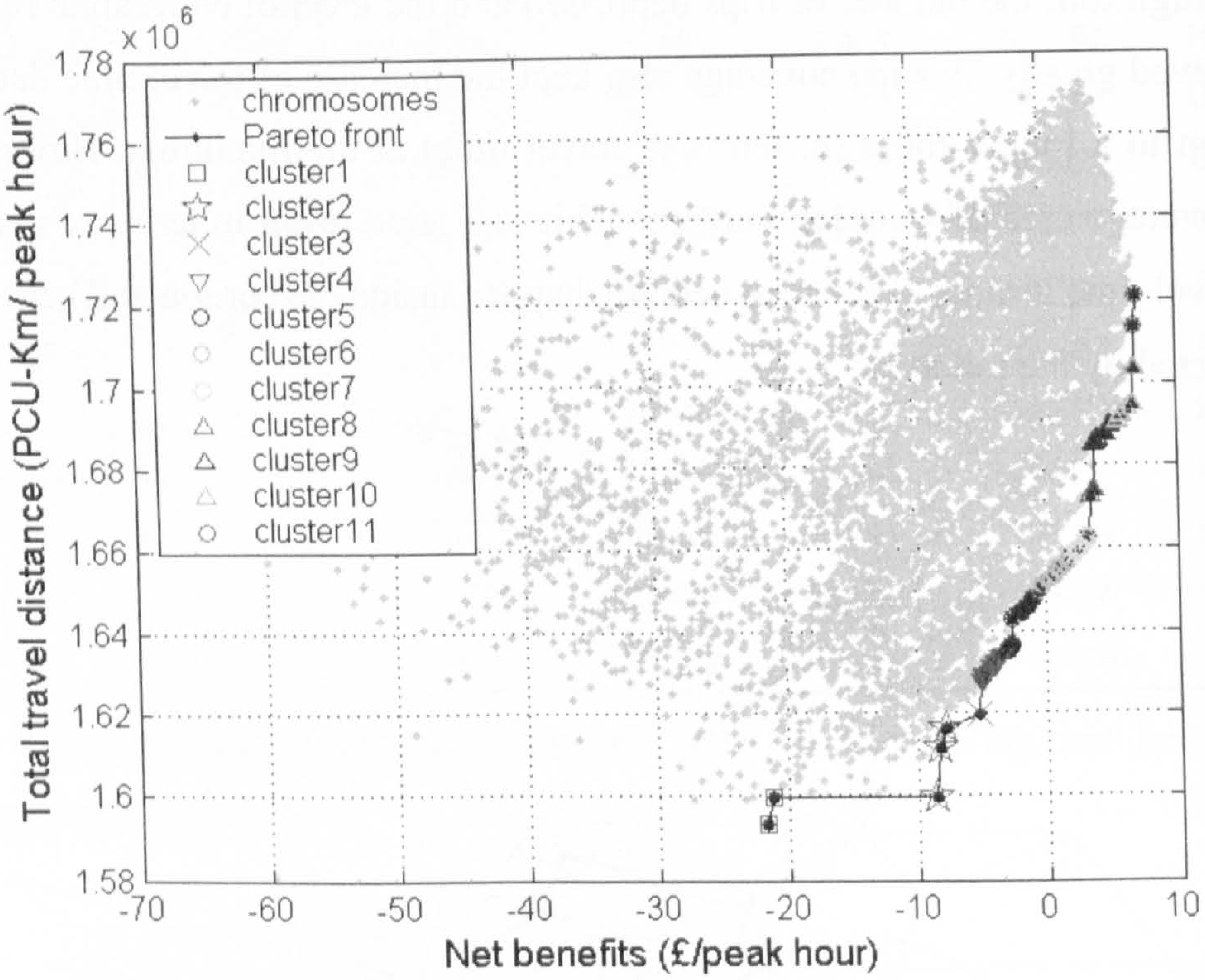


Figure 8-16 Pareto solutions and corresponding Pareto front for MOOP problem with objectives of maximising net benefits and minimising total travel distance

Cluster	No. of members	Members	Min. TTDis	Max. TTDis	Min net benefit	Max net benefit
1	2	WD1-WD2	1593263	1599510	-21668.0	-21173.9
2	3	WD3-WD5	1599614	1616113	-8674.60	-7905.55
3	1	WD6	1619556	1619556	-5211.29	-5211.29
4	17	WD7-WD23	1627963	1634462	-5142.42	-3473.19
5	13	WD24-WD36	1635282	1648079	-3110.65	-983.71
6	16	WD41-WD55	1649666	1658134	-560.73	1799.46
7	11	WD56-WD66	1658195	1662646	2136.83	3383.14
8	3	WD67-W69	1672292	1673434	3425.89	3753.28
9	10	WD70-WD79	1684786	1689555	3792.03	5066.51
10	14	WD80-WD93	1690398	1703384	5619.87	6999.95
11	2	WD94-WD95	1713466	1720864	7037.11	7207.76

Table 8-5 Clusters of the non-dominated solutions (maximizie net benefit and minimise total travel distance)

Note that the optimal uniform toll levels found for all non-dominated solutions increase as the total travel distance decreases. Also some relationship between the amount of demand left in the network and the total travel distance is also found. Figure 8-17 plots the total travel distance against the total demand for all non-dominated solutions from the fourth test.

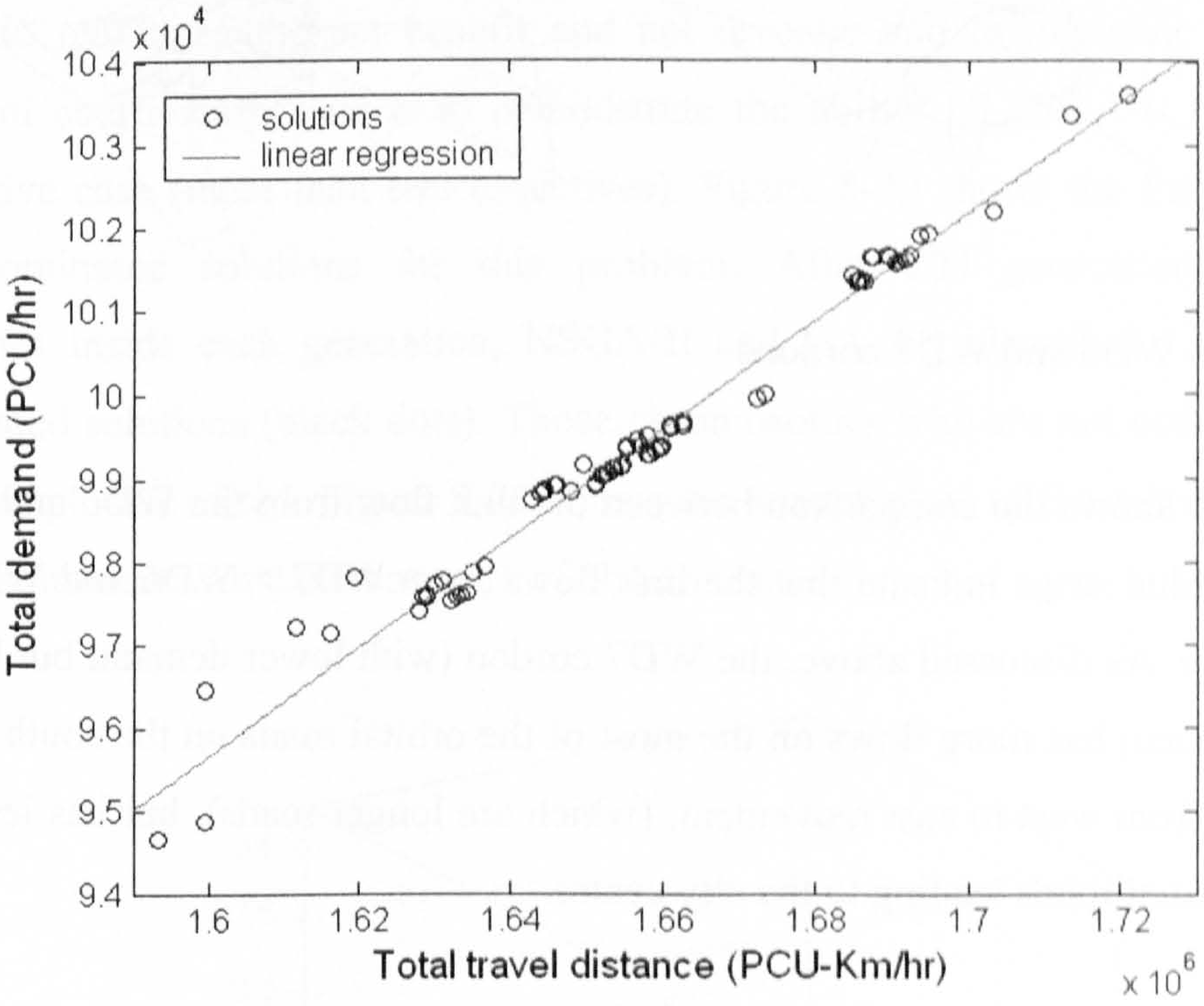


Figure 8-17 Relationship between the total travel distances and total demand in the network

From the figure, there seems to be a highly linear relationship between the amount of traffic in the network and the total travel distance. The traffic volume constitutes the major part of the objective function of the total travel distance (flow times distance). However, from the plot some of the cases even with a lower demand may generate a higher total travel distance. This is mainly due to the diversion of the traffic to a longer route.

Figure 8-18 shows the locations of the WD6 and WD7 cordons. The total travel distance for the WD6 and WD7 are 1619556 and 1627963 PCU-Km/hr respectively and the total travel demand in the network are 97828 and 97413 PCU/hr in that order. Despite a lower demand, WD7 generated a slightly higher total travel distance.

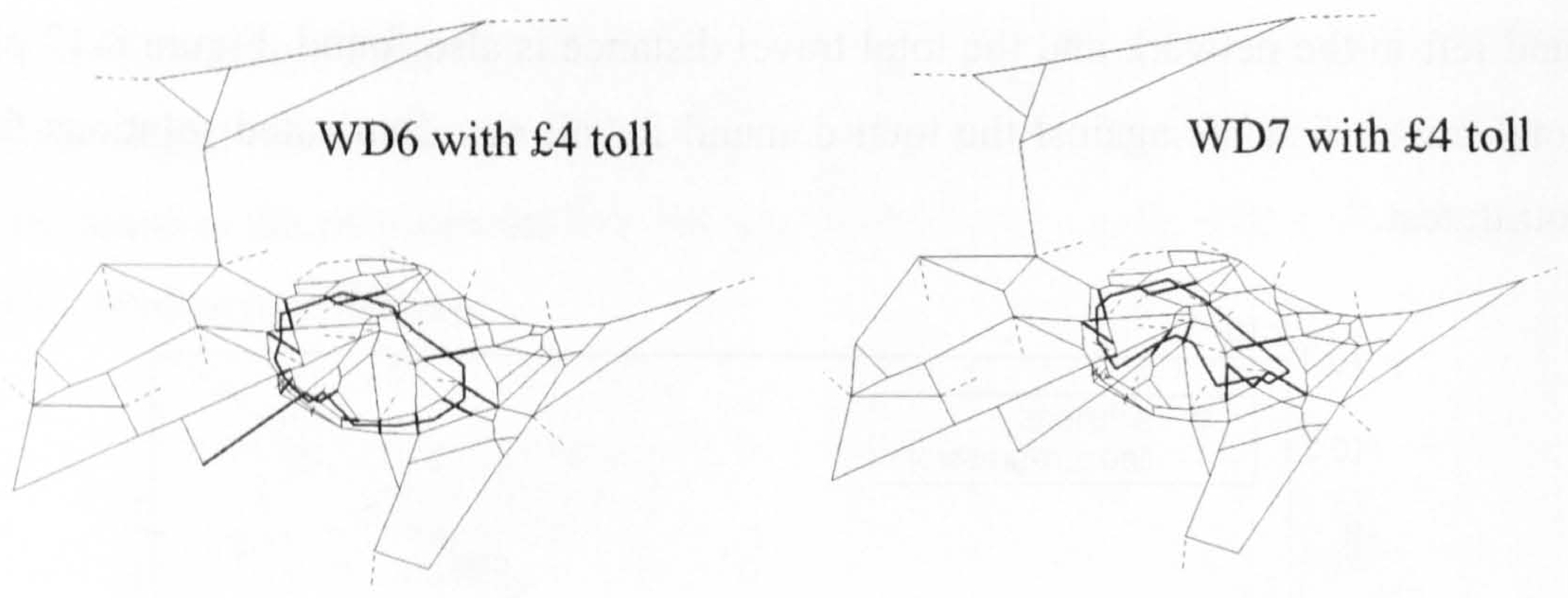


Figure 8-18 WD6 and WD7 cordons

Figure 8-19 shows the comparison between the link flow from the WD6 and WD7. The green and blue strips indicate that the link flows from WD7 > WD6 and WD7 < WD6 respectively. As discussed above, the WD7 cordon (with lower demand but higher total travel distance) has more flows on the most of the orbital roads on the south of the city, especially from west to east movement, (which are longer roads), but has less flows on the main radial roads leading to the city centre.

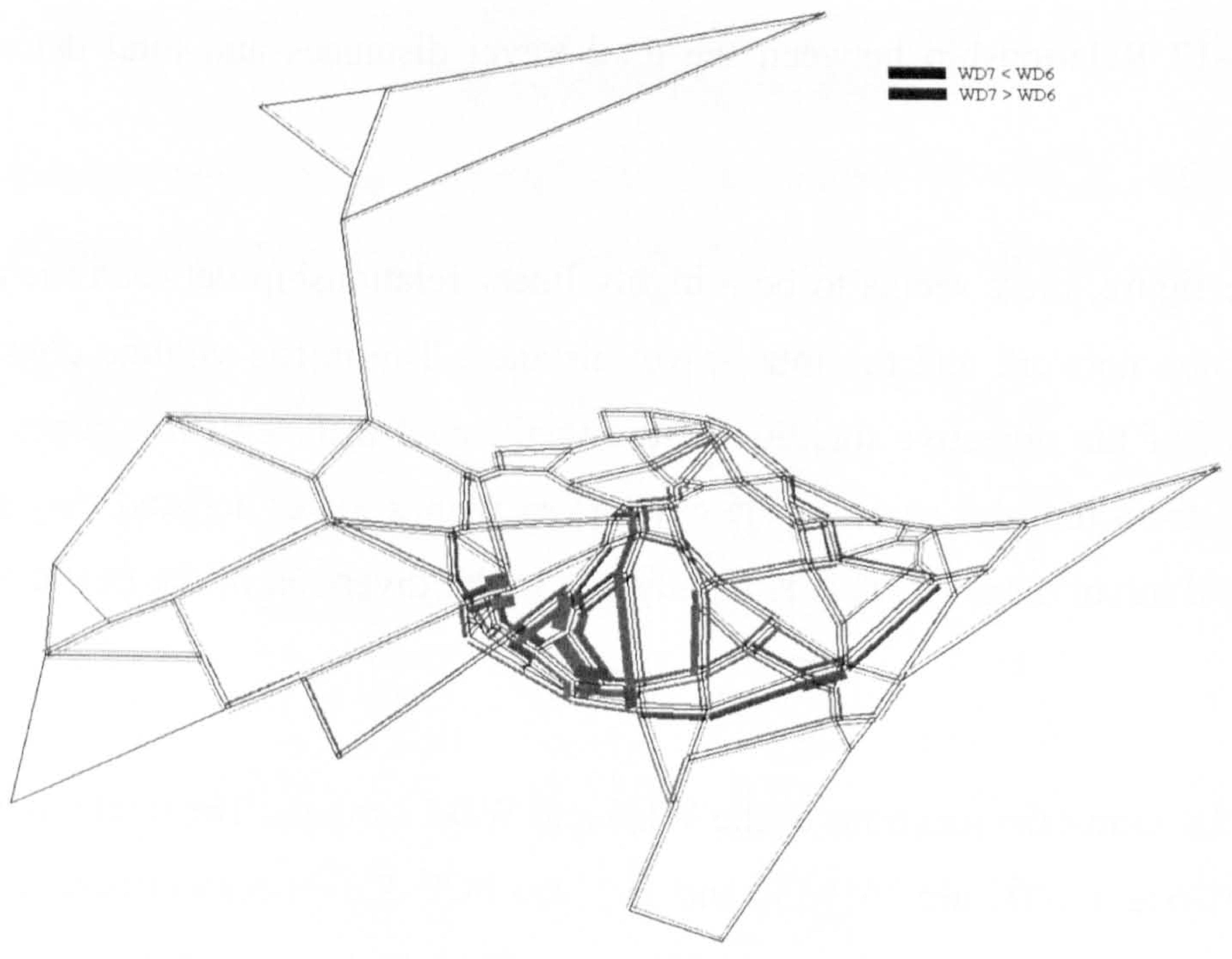


Figure 8-19 Comparison of link flows between WD6 and WD7

(v) Maximising net benefit, Maximising net revenues, and Minimising equity impact

This section presents the result for the last test with the NSGA-II with posterior preference articulation. Three objectives are considered in the test including the objectives to maximise the net benefit and net revenue and to minimise the equity impact (Gini coefficient). This is to demonstrate the ability of NSGA-II to handle a multiobjective case (more than two objectives). Figure 8-20 shows the Pareto surface and non-dominated solutions for this problem. After 200 generations with 50 chromosomes inside each generation, NSGA-II and GA-AS identified 128 different non-dominated solutions (black dots). Those chromosomes who are not non-dominated solutions are plotted in red. Figure 8-21 shows the scatter plots of the whole 128 non-dominated solutions with different pairs of objectives.

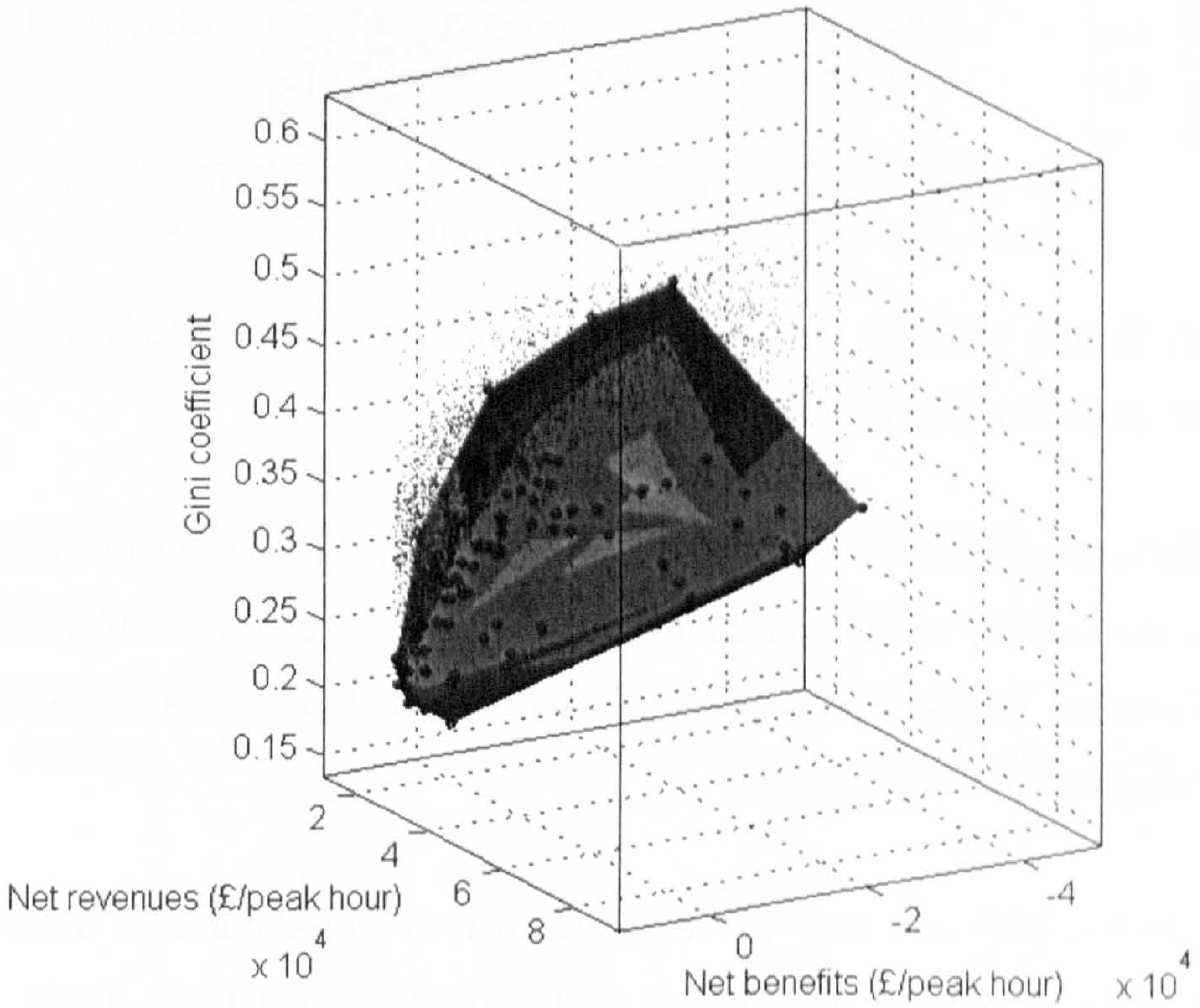


Figure 8-20 Pareto surface and non-dominated solutions for the MOOP-CD with three objectives (net benefit, net revenue, and equity)

From the whole 128 non-dominated solutions, 13 are non-dominated solutions for the pair of the objectives of maximising the net benefit and net revenue, 19 are non-dominated solution for the pair of the objectives of maximising the net benefit and minimising the Gini coefficient, and 10 are non-dominated solution for the pair of the objective of maximising the net revenue and minimising the Gini coefficient. These pair-wise non-dominated solutions from the set of non-dominated for the three objective problem are plotted in Figure 8-21.

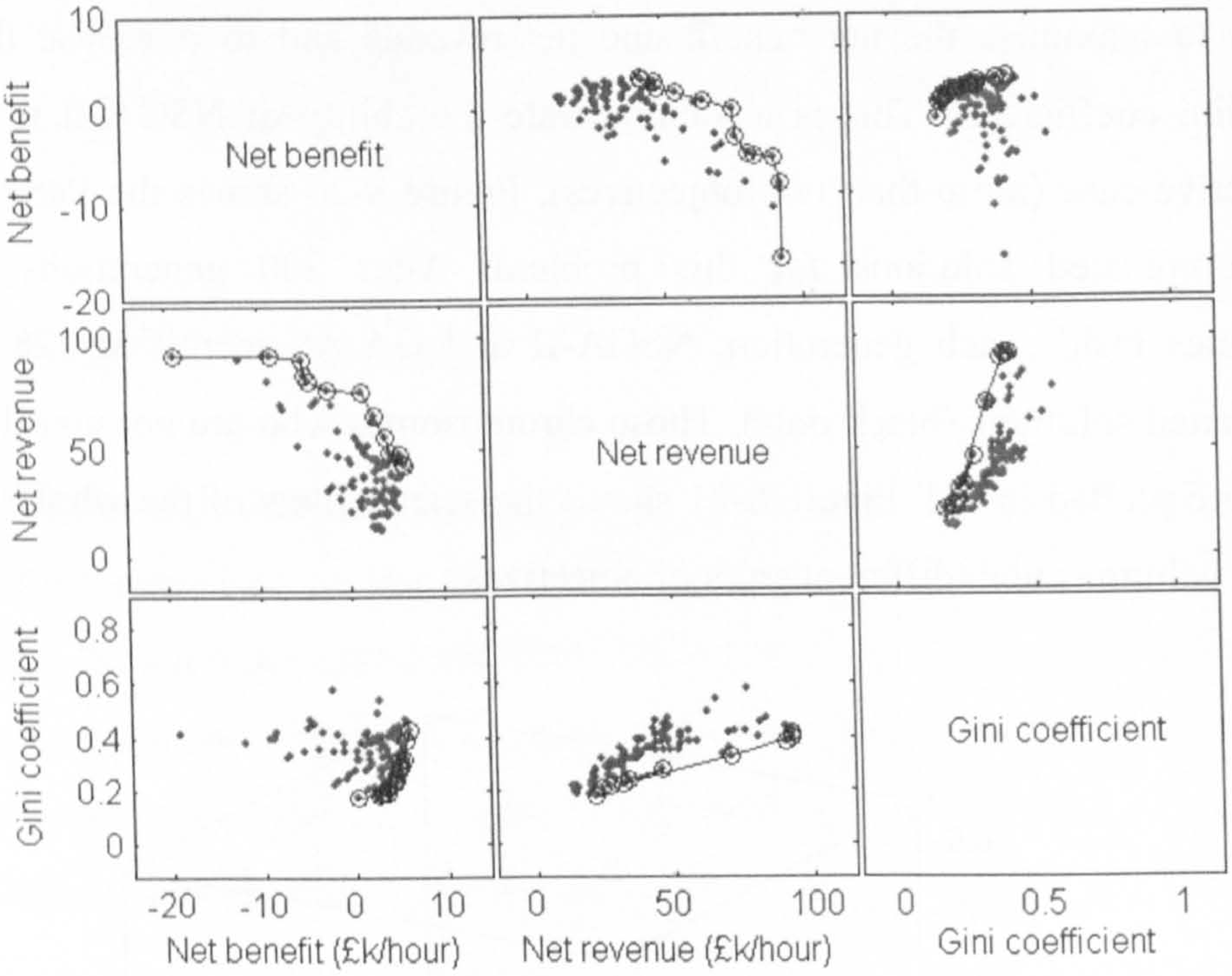


Figure 8-21 Scatter plots for the MOOP-CD results with three objectives (net benefit, net revenue, and Gini coefficient)

Clearly, the total number of pair-wise non-dominated solutions (42 in total) is much lower than the number of the non-dominated solutions for the three objective problem (128). This means adding more objectives into the MOOP increases the number of the possible non-dominated solutions.

Similarly to the previous test, the high number of non-dominated solutions (128 solutions) makes the detail presentation of the results impractical. Again, 128 non-dominated solutions are clustered into 15 different groups according to the levels of the objectives. These groups are depicted in Figure 8-22 with different plots. The detail of each cluster is given in Table 8-6. As mentioned earlier, in the real decision process, the planner may be presented with the information as shown in Table 8-6 to prevent overloading the information to the planner. Then, he or she can choose to look into the detail of some of the cluster in order to make the final decision on the best compromise cordon scheme.

An alternative to this approach is to allow the decision maker to interact dynamically with the optimisation process in order to focus and allocate more computational effort on the area of interest on the Pareto surface/front. This method is the progressive preference articulation approach which will be tested in the following section.

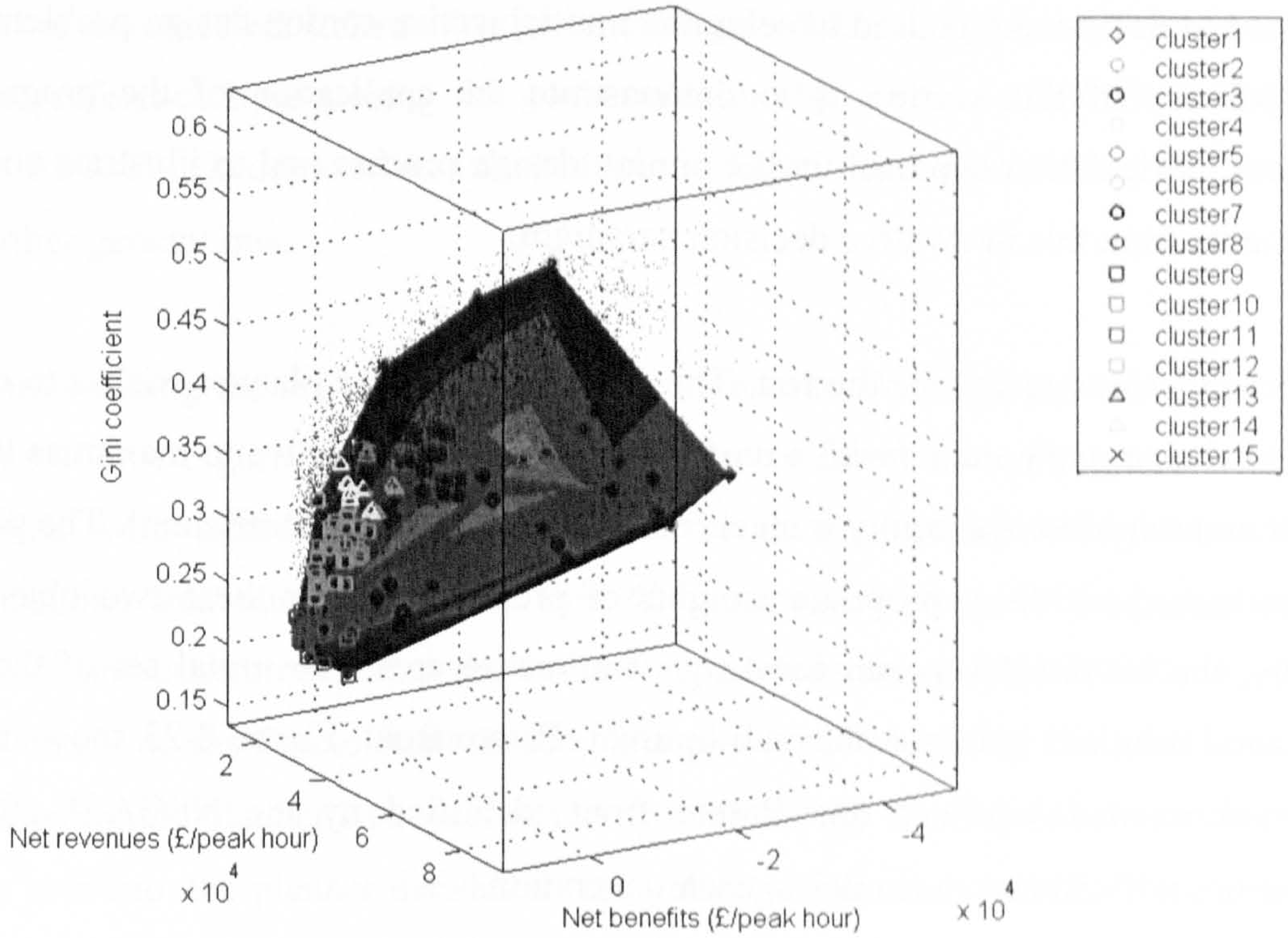


Figure 8-22 Pareto surface and 15 clusters of the non-dominated solutions for the MOOP-CD with three objectives (net benefit, net revenue, and equity)

Cluster	No. of members	Min net benefit (£)	Max net benefit (£)	Min net revenue (£)	Max net revenue (£)	Min Gini	Max Gini
1	1	-19021.0	-	92508	-	0.410	-
2	4	-11992.0	-8661.2	80472	92245	0.378	0.422
3	3	-7454.9	-5668.1	51853	70159	0.328	0.35
4	6	-5134.0	-1822	69430	90003	0.357	0.467
5	3	-2423.1	-1111.5	46445	58184	0.352	0.383
6	1	-2366.9	-	76367	-	0.572	-
7	4	-2329.1	-767.07	27651	33129	0.228	0.262
8	7	-26.8	1554.6	32661	45564	0.280	0.353
9	20	58.3	4048.5	13039	26405	0.175	0.230
10	13	562.3	3978.7	42958	55056	0.373	0.421
11	31	1162.9	5302.2	20241	32062	0.241	0.329
12	1	1423.1	-	75561	-	0.421	-
13	2	2397.8	2773.3	61922	65269	0.487	0.536
14	19	2621.2	5486.7	29913	41926	0.327	0.391
15	13	4378.0	5900.2	42757	48120	0.413	0.464

Table 8-6 Clusters of the non-dominated solutions (maximizie net benefit and net revenue and minimise Gini coefficient)

8.6.3 Progressive preference articulation case

In Sections 8.6.1 and 8.6.2, the methods of priori and posterior preference articulation were applied to the MOOP-CD. In this section, the progressive preference articulation approach is adopted to solve the MOOP-CD. The extension of the NSGA-II method as described in Section 8.5 is used to solve the multiobjective cordon design problem. The main purpose of this section is to demonstrate the application of the progressive preference articulation approach in the cordon design process and to illustrate some of the benefits from this innovative decision paradigm.

A scenario is constructed for the test. The scenario is that the planner wishes to design an optimal charging cordon (with a uniform toll) with the objective to maximise the net benefit and minimise the equity impact (measured by the Gini coefficient). The planner is uncertain about the appropriate weights of preferences over these two objectives. Initially, the NSGA-II was run for 40 generations to create an initial set of the non-dominated solutions and (possible sub-optimal) Pareto front. Figure 8-23 shows the set of non-dominated solutions and Pareto front identified by the NSGA-II after 40 generations with 50 chromosomes in each generation.

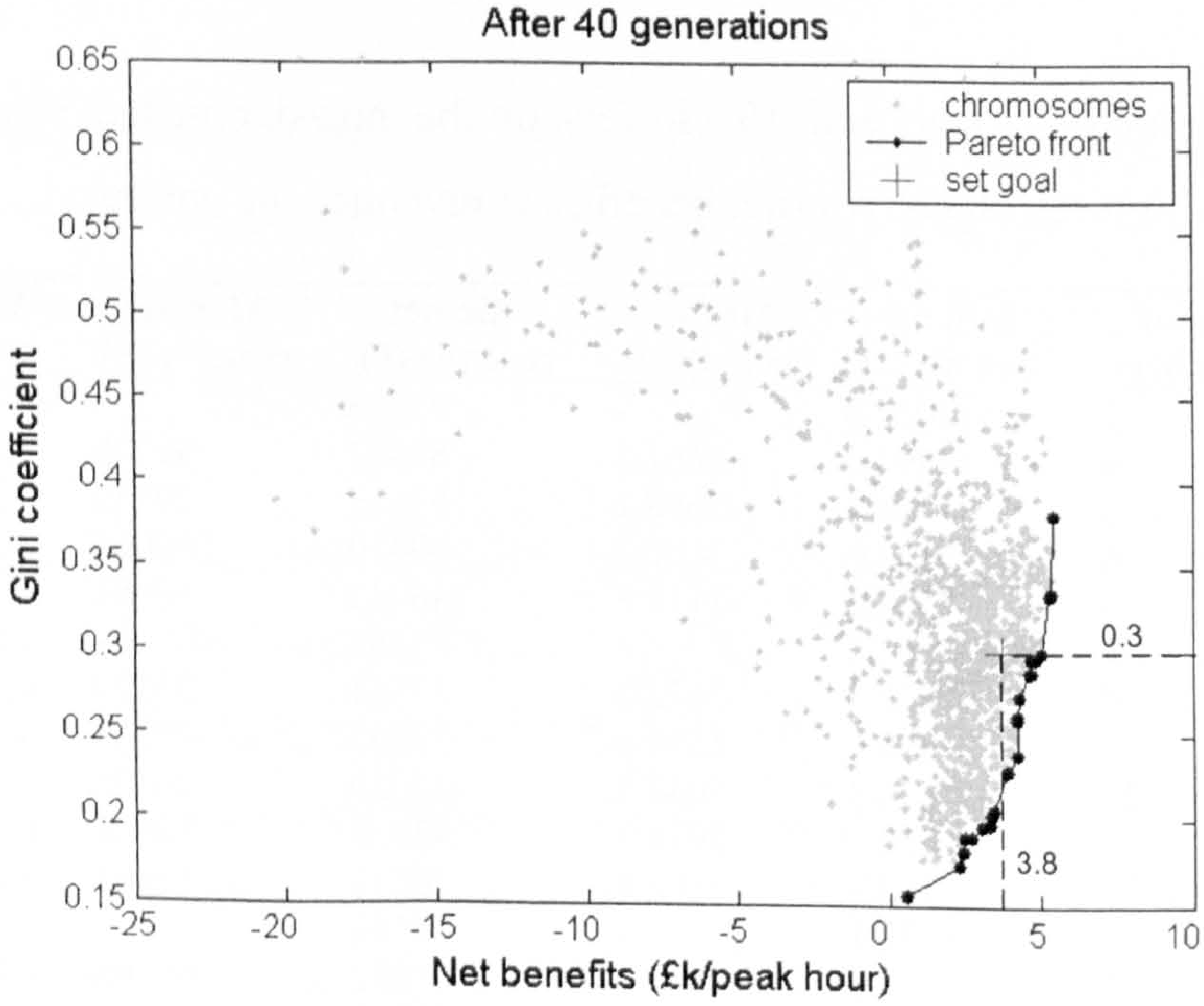


Figure 8-23 Pareto front and progressive goals set after 40 generations of NSGA-II

As shown in Example 8-4, the decision maker (transport planner in this case) can exploit the given initial information on the possible trade-off between two objectives (shown by the Pareto front found after 40 generations) to define the area of interest on the Pareto front presented. The planner will provide the updated preference vector to guide the NSGA-II to focus on the main trade-off area of interest. In this scenario, after observing the trade-off information given in Figure 8-23, the planner is assumed to set his/her preference by imposing the goals on both objectives, net benefit > £3815 and Gini < 0.30. This is depicted as a dashed line in Figure 8-23 representing the focus area for further generations.

After introducing the set goals for both objectives, the NSGA-II proceeded for further 40 generations. Figure 8-24 shows the found Pareto fronts after 80 generations where the goals were introduced for both objectives during the last 40 generation. The chromosomes generated during the first 40 generations are plotted in grey and the chromosomes generated during the last 40 generations are plotted in gold. Notice the advancement on the Pareto front on the particular area focused by the goals set earlier. At this point, the planner can observe the updated Pareto front again and set new goals. In this scenario, the planner decided to set a new goal for the net benefit (net benefit > £4450) and keep the same goal for the Gini coefficient. This new goal is plotted as a dashed line in Figure 8-24.

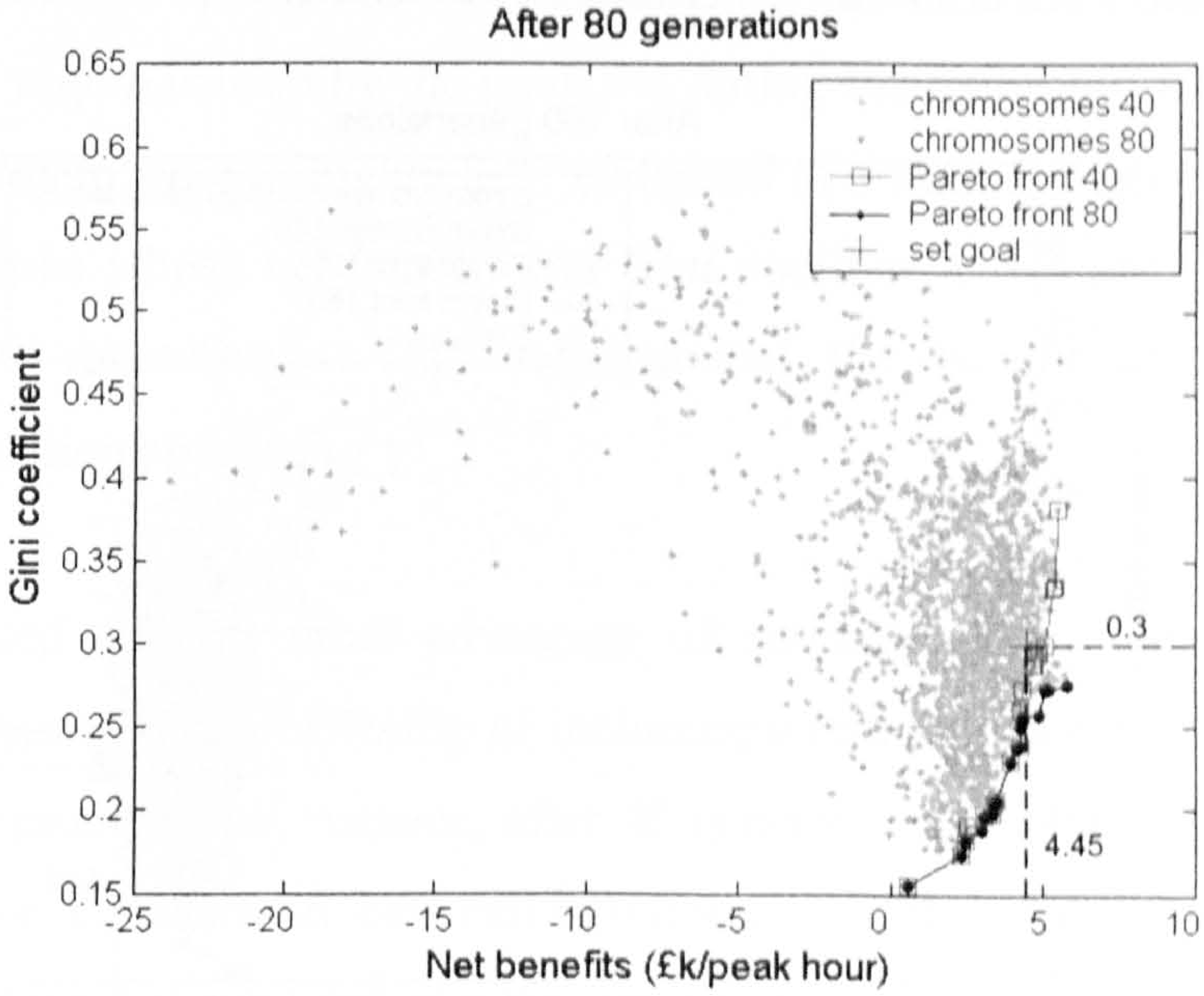


Figure 8-24 Pareto front and progressive goals set after 80 generations of NSGA-II

Then, the algorithm is run for a further 80 generations. Figure 8-25 and Figure 8-26 show the Pareto front found after 120 and 160 generations respectively. Note that the goals were revised after 160 generations. The goal for the net benefit was set to be greater than £5086 and the goal for the Gini coefficient was set to be less than 0.28 limiting further the area of interest (dash lines in Figure 8-26). Again, after setting the new goals the algorithm was let to run until the end of the whole process (200 generations). Figure 8-27 shows the final Pareto front found after 200 generations of NSGA-II.

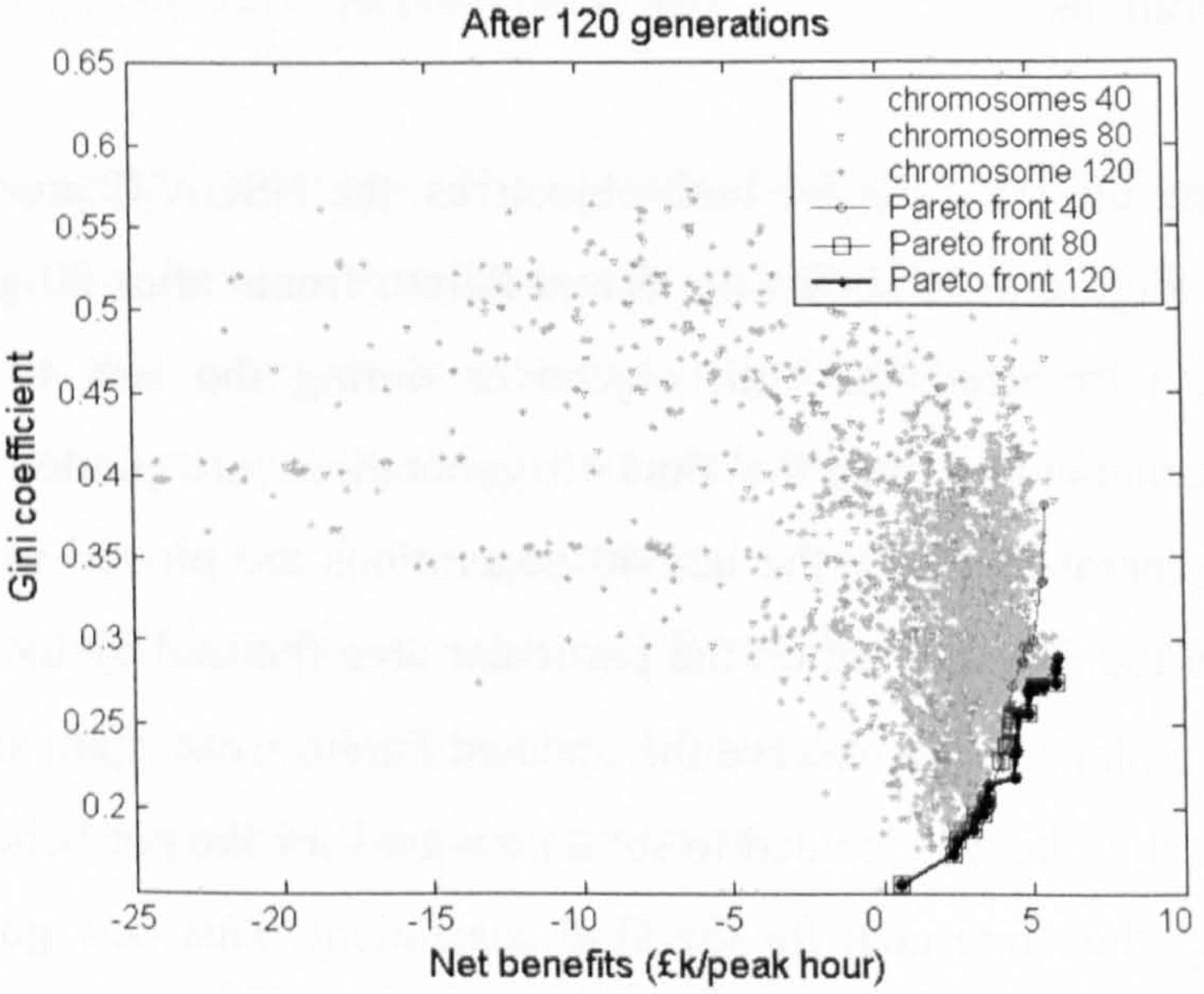


Figure 8-25 Pareto front after 120 generations of NSGA-II

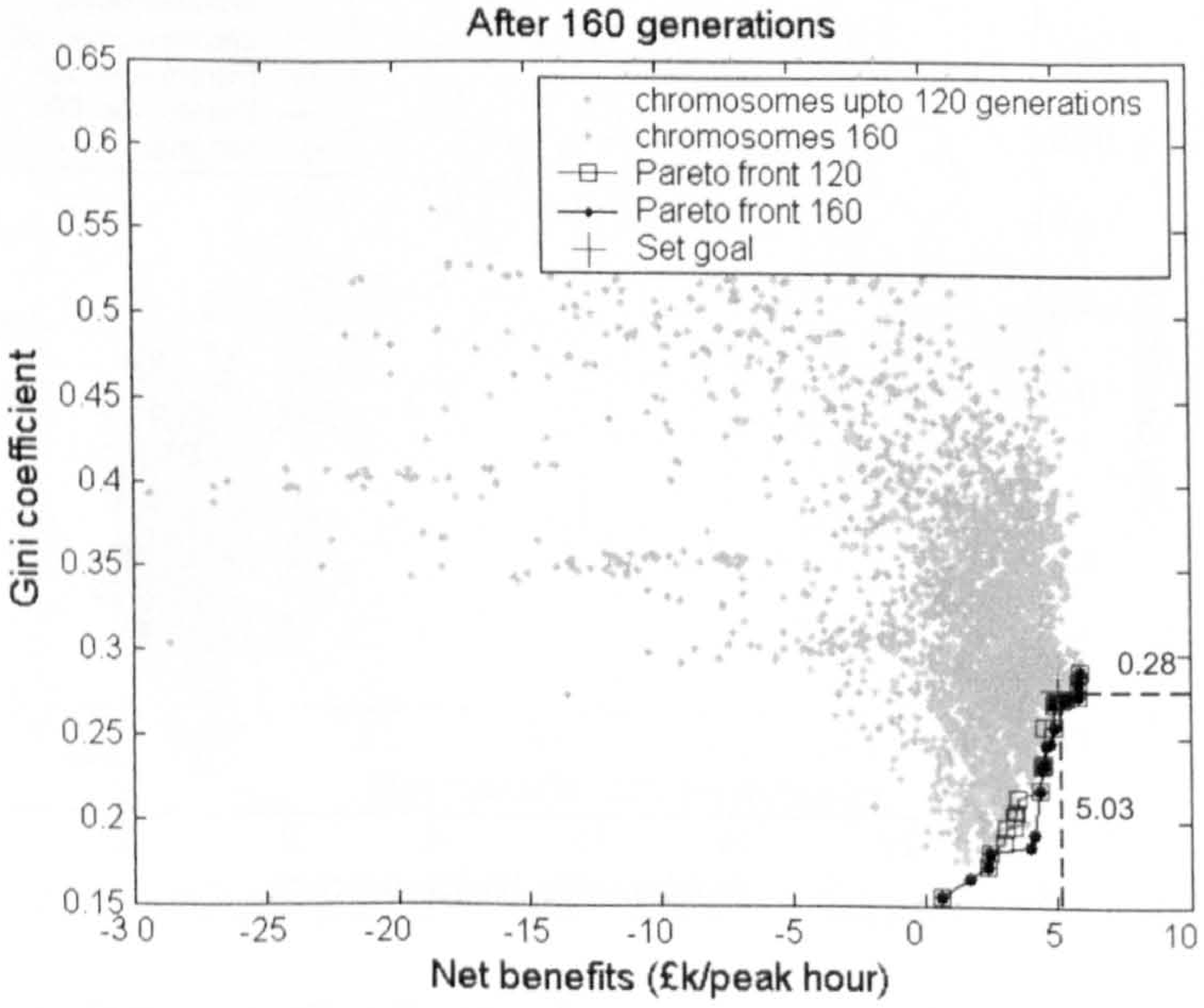


Figure 8-26 Pareto front and progressive goals set after 160 generations of NSGA-II

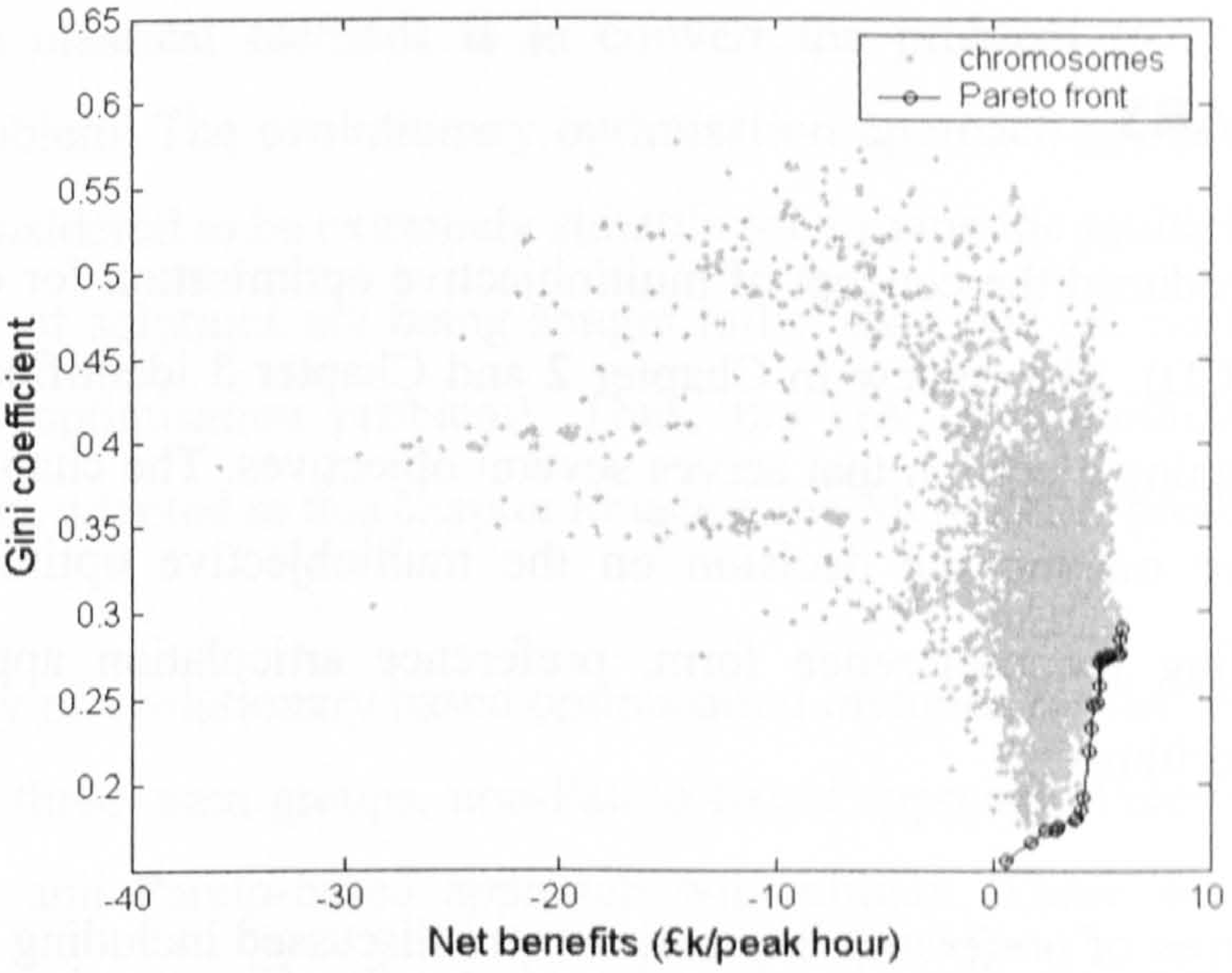


Figure 8-27 Pareto front from the progressive preference articulation after 200 generations

29 non-dominated solutions were found. From this set, 20 of them satisfied the goals set after the first 40 generations (net benefit > £3815 and Gini < 0.30). This Pareto front, generated under the progressive preference articulation process, clearly covers a narrower range of the non-dominated solutions compared to the result from the posterior preference articulation approach shown earlier in Figure 8-11. This is due to the restricted search region caused by the goals set during the optimisation process. On the other hand, the Pareto front in Figure 8-27 consisted of a slightly higher number of non-dominated solutions whose net benefit and *Gini* coefficient is higher than £3815 and lower than 0.30 respectively. This demonstrated the benefit of the progressive preference articulation approach.

It should be noted that the other advantage of adopting the progressive preference articulation approach is the possibility of including additional design constraints during the optimisation process. For instance, after 40 generations of NSGA-II, in addition to the observation on the trade-off between two objectives, the planner may also analyse the outcome in terms of total travel time, total travel distance, net revenue, etc. The planner then can also decide to introduce additional constraints into the optimisation problem afterward based on this observation. The constraints can be introduced within the preference vector framework as explained earlier in Section 8.5.1.

8.7 SUMMARY

This chapter introduced the concept of multiobjective optimisation for charging cordon design (MOOP-CD). The review in Chapter 2 and Chapter 3 identified the need for a method for designing a cordon that serves several objectives. The chapter reviewed the key elements for making the decision on the multiobjective optimisation problem (MOOP) including the preference form, preference articulation approach, and the optimisation algorithms.

Four different types of preference expression were discussed including weight, priority, goal/target, and linguistic ranking. The weighting approach (or utility based approach) is probably the most widely adopted method due to its simplicity. The linguistic ranking approach is the most recent approach which becomes one of the feasible approaches following the development in the area of fuzzy reasoning and evolutionary optimisation.

Regardless of the form of the preference, different stages of preference articulation can be adopted. Three possible approaches for deciding on the preference were discussed in the chapter including the priori, posterior, and progressive preference articulation approaches. The combination between the priori preference articulation and the weighting or utility based approach is probably the simplest method for solving the multiobjective problem. This is also the case in transport area. However, several drawbacks of this approach were discussed in the chapter. The main problem seems to be the difficulty in making the judgement on the preference prior to obtaining the actual trade-off information between different objectives. The recent development of evolutionary optimisation or in particular genetic algorithm allows a better treatment of the multiobjective optimisation problem.

Before presenting the actual algorithm for solving the multiobjective problem, the chapter reviewed different possible methods both from classical optimisation and evolutionary optimisation paradigms. Although the classical optimisation method may be adopted to solve the problem, it was remarked that the nature of the single objective optimisation problem is significantly different from the multiobjective problem. This makes the classical optimisation method inappropriate due to the fact that the main strategy for the classical methods is to convert the problem to a single objective optimisation problem. The evolutionary optimisation approach operates on population basis. This is considered to be extremely suitable for solving the multiobjective problem where a number of solutions are being sought rather than one (as contrast to a normal single objective optimisation problem). Thus, the GA or evolutionary optimisation based method was adopted in this chapter to tackle the MOOP-CD problem.

There is a variety of evolutionary based optimisation methods for the MOOP that can be categorised into three main groups, non-Pareto based approach, Pareto-based approach without elitism, and Pareto-based approach with elitism. These were reviewed and discussed in the chapter. The Pareto-based approach with the elitism strategy was considered to be the most up to date and efficient algorithm. The method of NSGA-II proposed by Deb *et al* (2000) was integrated with the GA-AS method developed in Chapter 6 for tackling the MOOP-CD in this chapter.

Two variations of the NSGA-II methods were proposed. The first one is the standard approach based on the posterior preference articulation decision paradigm. The NSGA-II algorithm is used to identify a number of non-dominated solutions in which the decision maker eventually selects the best compromise solution from these non-dominated solutions (without having to express his/her preference for different objectives in advance). The second version of the NSGA-II is the progressive preference articulation based approach. The framework for allowing the interaction between the decision maker and the optimisation algorithm was presented in the chapter. The main modification was the definition of the non-dominated solution in which the idea of preference vector used to compare to solutions was proposed. With this new definition of non-dominated solutions, the user can sequentially set the goals/targets for different objectives during the optimisation process in order to guide the search to the area of his/her interest.

Three different set of tests were conducted with the Edinburgh network. The first set was for the priori preference articulation decision paradigm. Four tests were conducted (two tests with two objectives and two tests with three objectives). Each of them received different weights for different objectives that were used to create the weighted sum objective function. The GA-AS approach was then applied to solve all four problems with the weighted sum objective (hence the problems became normal single objective optimisation problems). The remark was made regarding the difficulty in setting appropriate weights due to the scaling problem with different objective functions.

The second set of the tests was concerned with the posterior preference articulation paradigm. Four tests with two objectives and one test with three objectives were the objects of the methodological experiment. The NSGA-II with GA-AS was successfully applied to all problem to generate the set of non-dominated solutions. The implications of the results from the policy perspective were briefly discussed. This issue will be dealt with in the following chapter. The results revealed a possible weakness of the NSGA-II in which the set of non-dominated solutions identified by the algorithm may not be well spread over the Pareto front. Finally, the last test was for the progressive preference articulation approach. A scenario was set up to demonstrate the application of the NSGA-II with the progressive preference articulation.

Overall, the chapter successfully developed innovative optimisation approaches for dealing with the MOOP-CD problem. These two methods, posterior and progressive preference articulations, offer a new way for dealing with the selection and the design process of the transport policy, in particular the charging cordon scheme design. It is difficult to judge the full advantage of these new methods at this stage. A fuller experiment with a more realistic case and participant is required to fully evaluate the additional advantage from these innovative methods compared to a more traditional one.

CHAPTER 9 POLICY IMPLICATIONS

If we know what it was we were doing it would not be called research, would it? Albert Einstein

9.1 INTRODUCTION

The previous four chapters (Chapters 3-8) discuss separately a number of possible approaches for designing a road pricing scheme, e.g. judgmental design, theoretical design, or constrained design. Obviously different design philosophies yield different scheme designs and performances. In this chapter, all numerical results with the Edinburgh network from the previous chapters are revisited and used in the discussion on some of their key implications on the charging cordon design. Some of these points are discussed with the reference to the real scheme design of the Edinburgh congestion charging scheme (the scheme is planned to start in 2006) or the extension of the London congestion charging scheme currently being discussed at the time of writing this thesis.

This chapter is structured into three further sections. The next section gathers the numerical results from different tests conducted previously in order to analyse the effect of the toll location and toll level on the performance of the scheme. This section also looks at the trade-off between the practical constraints and the benefit of the scheme, and the issue of multiobjective design. Then, Section 9.3 discusses the implication of this research to real-world decisions on road pricing scheme design. In particular, the section explains the extent to which the development of different algorithms in this thesis may support the decision maker in the real world. Finally, the last section concludes the chapter.

9.2 PERFORMANCE COMPARISON

9.2.1 Effect of toll levels and locations

As described in the objective and motivation of this research, it is of practical and theoretical interests to analyse the extent to which the toll level and location affect the performance of the road pricing scheme. The term "performance" discussed here mainly concentrates on the net benefit of the scheme (following the Marshallian measure of social welfare improvement as detailed in Chapter 4). Firstly, the effect of the toll level on the performance of a road pricing scheme will be discussed.

There are several tests in the previous chapters that can be used to analyse the effect of the toll level on the net benefit of the scheme. In Chapter 4, three judgmental cordons were identified based on the judgmental design criteria specified in Chapter 3. In addition, the optimal uniform toll and varied tolls (using GA-CHARGE) for each of these judgmental cordons were found in Chapter 4. In Chapter 5, GA-AS was used to find the optimal cordon with uniform toll, OPC1 cordon. Similarly, GA-ASII was used in Chapter 4 to find the optimal double cordon, D-OPC. All cordons were optimised with the objective of maximising the net benefit (see its definition in Chapter 4).

In addition to the aforementioned results, the top-15 links with highest marginal social cost tolls (from the system optimum) are included into the comparison. This is mainly to provide an estimate of the benefit of the second-best tolling scheme. The system optimum assignment (with elastic demand) was carried out with the Edinburgh network. The by-product of this assignment is the marginal cost tolls for all links in the network. The top-15 links are chosen from those links with the highest marginal cost tolls. Figure 9-1 shows these 15 links with the label of the ranking of these links. Table 9-1 shows the marginal social cost toll for these 15 links from the system optimal assignment. An optimal uniform toll for the top-15 is found to be £0.75. GA-CHARGE is also used to optimise the tolls for the top-15 charging system.

Table 9-2 presents the test results for the inner cordon 1, inner cordon 2, outer cordon, OPC1, Top-15, and D-OPC (optimal double cordon). From the results in Table 9-2 Comparison of the performance of different charging regimes for the Edinburgh network, the effect of the complexity of the charge structure on the performance of the charging cordon scheme can be observed from the comparison of net benefit improvements from the inner1 cordon, outer cordon, and top-15 cordon. The variable tolls found by GA-CHARGE for the inner1 cordon, outer cordon, and Top-15 cordon generate the net EB improvement of around 50%, 70%, and 95% higher than those with optimal uniform tolls. Obviously, the more complex toll regime gives a higher degree of freedom for the scheme design through which it can achieve a higher benefit.

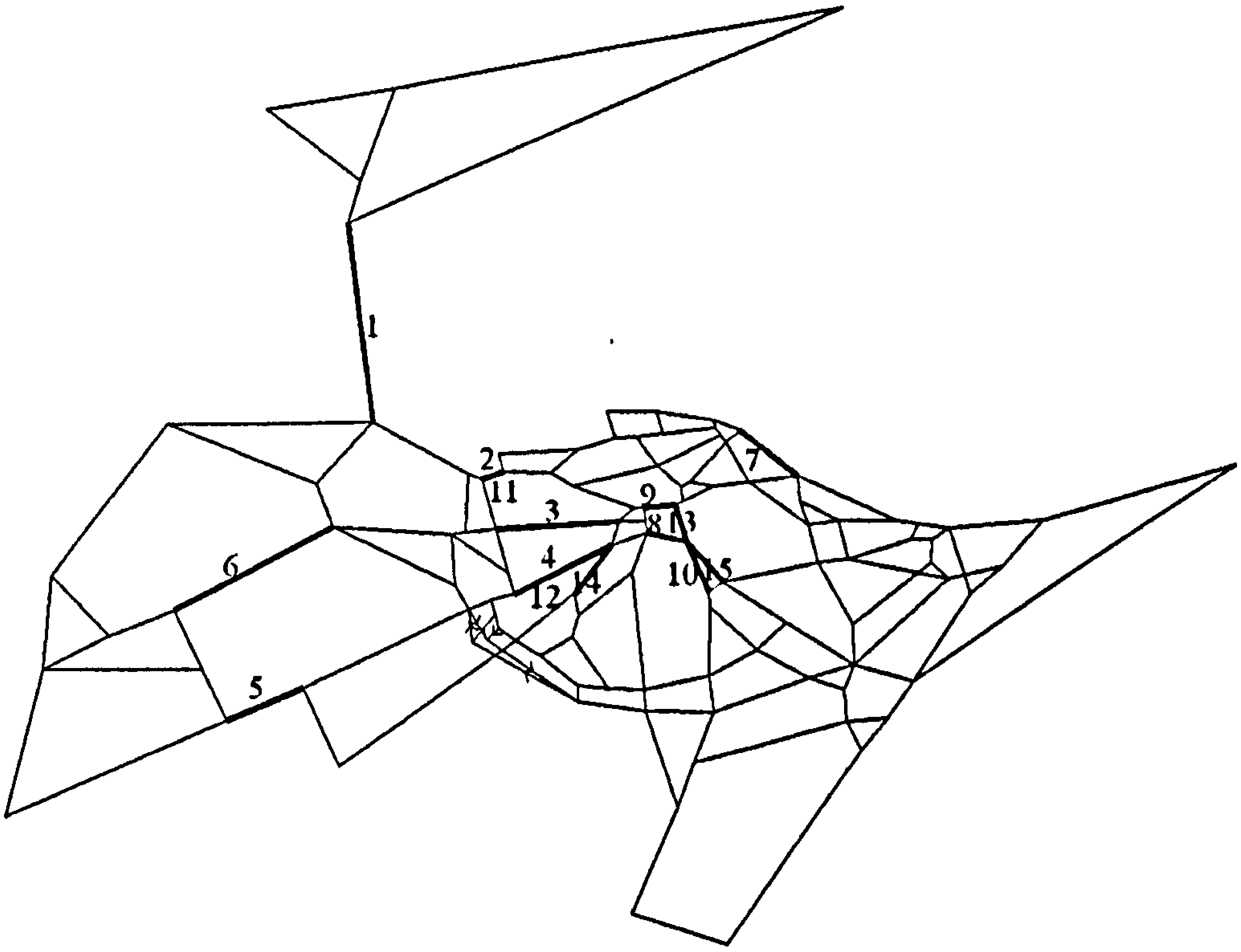


Figure 9-1 Top-15 links with the highest marginal social cost toll for the Edinburgh network (highlighted link with the label indicating the rank)

Rank	MC toll (£)	GA-CHARGE toll (£)	Rank	MC toll (£)	GA-CHARGE toll (£)
1	3.55	4.07	9	0.75	1.02
2	1.67	2.30	10	0.72	0.76
3	1.38	1.33	11	0.70	0.50
4	1.30	1.33	12	0.70	0.45
5	1.04	1.20	13	0.69	0.12
6	1.02	1.33	14	0.68	0.70
7	0.98	0.76	15	0.66	0.63
8	0.85	0.50			

Table 9-1 Marginal cost tolls from system optimum and GA-CHARGE tolls for the top-15 links of the Edinburgh network

Charging system	Optimal toll	Net EB (£k/hour)	Net Revenues (£k/hour)	TT. time (PCU-Hr/hr)	TT. distance (PCU-Km/hr)	Demand level (kPCU/hr)	Gini
Inner cordon 1	£0.50	2.10	11.70	56426	1767499	107.8	0.26
	varied	3.16	14.51	55993	1766306	107.6	0.32
Inner cordon 2	£0.75	3.99	3.99	55523	1767524	107.2	0.31
Outer cordon	£0.75	1.96	22.20	55505	1734647	106.4	0.20
	varied	3.31	22.87	55029	1733473	106.2	0.22
OPC1	£1.50	7.21	43.70	52325	1720864	103.6	0.41
Top-15	£0.75	9.21	39.24	52615	1730023	105.2	0.29
	varied	17.96	75.96	47919	1681672	103.9	0.56
D-OPC	£1.25	15.28	101.50	45157	1607555	96.8	0.48

Table 9-2 Comparison of the performance of different charging regimes for the Edinburgh network

The more complex toll structure can differentiate better the level of tolls imposed on different parts of the network. In most cases, the level of congestion and its relative marginal cost in the network vary between parts of the network due to the variation of the network capacity and demand. Figure 9-2 shows the delay on the links in the do-nothing scenario of the Edinburgh network. The width of the bandwidth represents the level of the delay on that link. From the figure, the links with very high level of delay are links A and B (which are also the links with the highest marginal cost). Obviously, there is a higher level of congestion on the west corridors of the city centre compared to other corridors. In general, the most congested area is the around the city centre inside the city by-pass. This modelling output is considered realistic since link A is the Forth Bridge which is a highly congested link (due to the lack of alternative route), and a high volume of demand travelling to the city centre comes from the west part of the Edinburgh region. Note that intuitively link A, the Forth Bridge, is the link with the highest toll and was included in the Top-15 links. This should also explain why the Top-15 links can generate a high level of net benefit improvement compared.

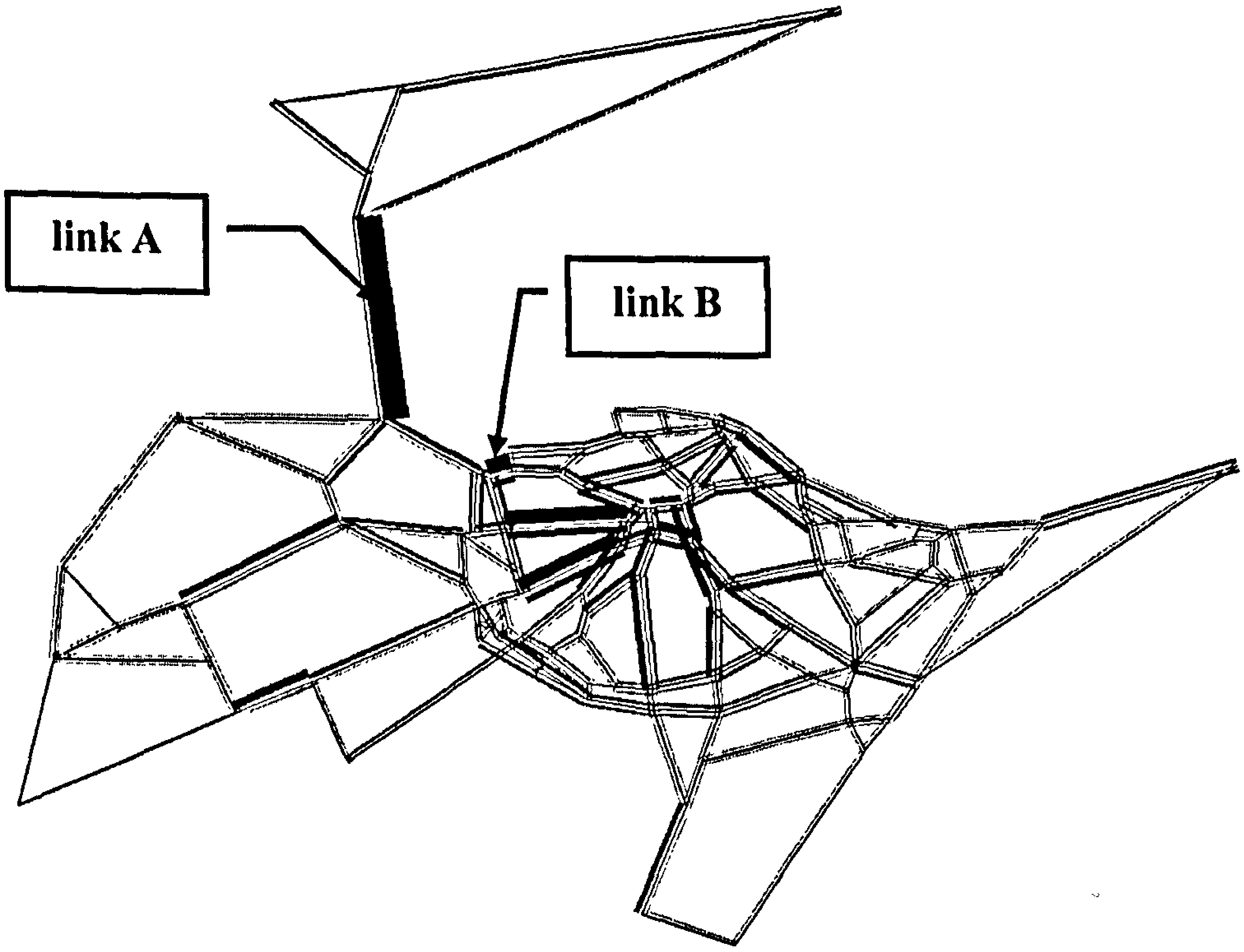


Figure 9-2 Link travel delay in the do-nothing scenario of the Edinburgh network (the width of the bandwidth represents the level of delay on that link)

Now consider the results of the outer cordon. Figure 9-3 compares the link delays between the test results of the outer cordon with the uniform toll and variable toll (delay in uniform toll – delay in variable toll cases). The width of the bandwidth represents the level of the difference in the link delay between the two cases. The black and grey bandwidths represent the delays in the variable toll case lower and higher than the uniform toll case respectively.

From Figure 9-2 and Figure 9-3, the variable toll for the outer cordon decreases delays more on links A and B. From Figure 5-25 (page 160) which shows the variable tolls for the outer cordon, GA-CHARGE determined a higher toll level in the north part of the charging cordon and near link B. This results in a reduction in the trips accessing the area inside the charging cordon in the northern part of the city centre. The reason for this adjustment of toll can be explained from Figure 9-4. Figure 9-4 plots the link flows in the network passing through links A and B in the do-nothing scenario. This figure represents the correlation of the contribution of the link flows from other links on the total flows on links A and B.

From Figure 9-4 and Figure 9-3, the higher reduction of the link delay in the variable toll scheme of the outer cordon compared to the uniform toll scheme occurred in the area/links related highly to links A and B (i.e. north-west part of the outer cordon). This partly explains the greater benefit of the variable toll compared to the uniform toll in which it can better adjust the toll level for different group of traffic according to maximise the social welfare improvement.

The other issue related to the toll level design is the benefit of adjusting the toll levels to their optimal values. With the Top-15 scheme, if the marginal cost tolls of these top 15 links (see Table 9-1) are implemented directly, the scheme yields the net benefit improvement of £16.50k/peak hour. With the toll levels found by GA-CHARGE, the Top-15 scheme generated the net benefit improvement of around £17.96k/peak hour which is about 8% higher than the benefit from the marginal social cost toll scheme.

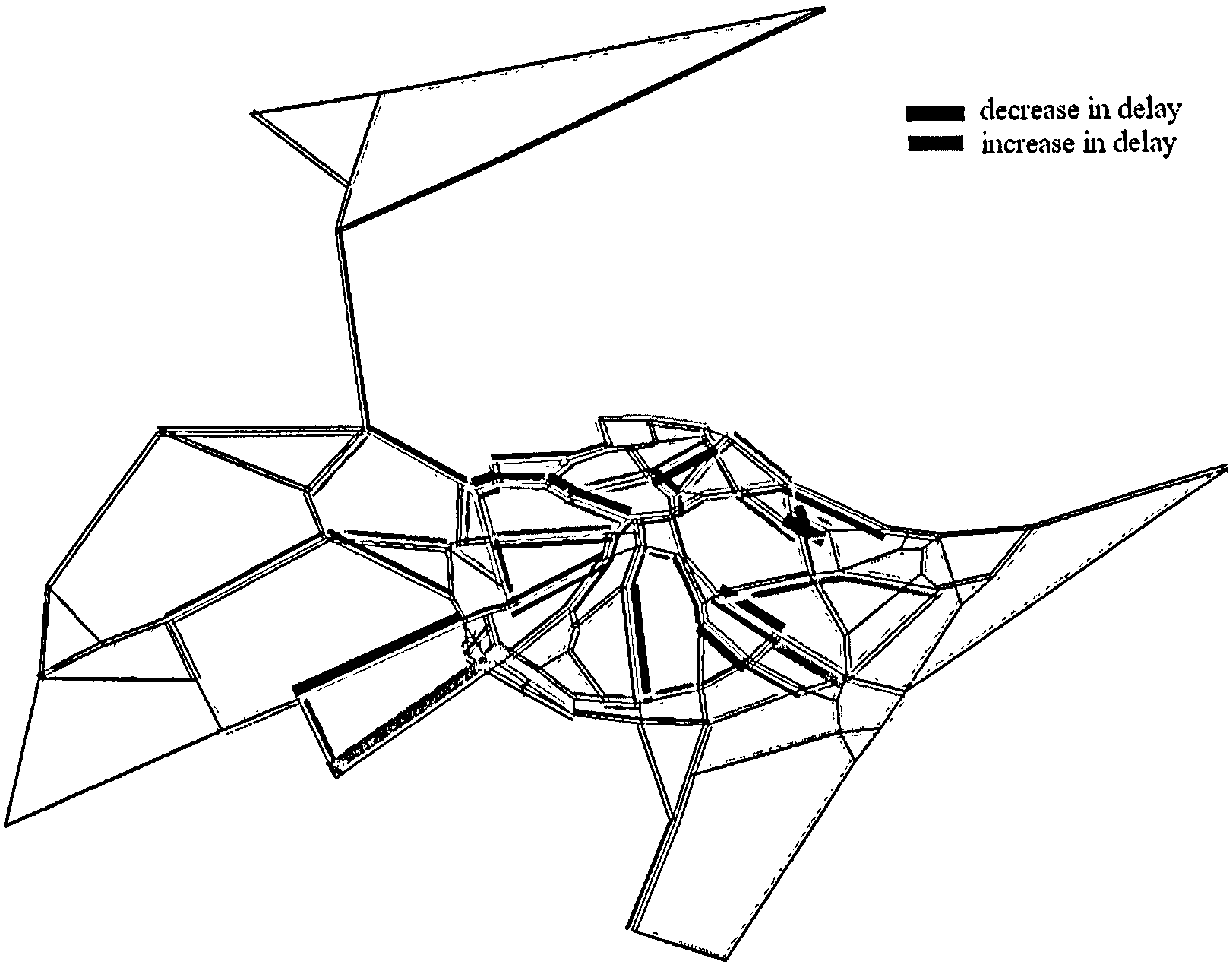


Figure 9-3 Comparison of the link delay for the outer cordon with the uniform toll and variable toll (uniform – variable)

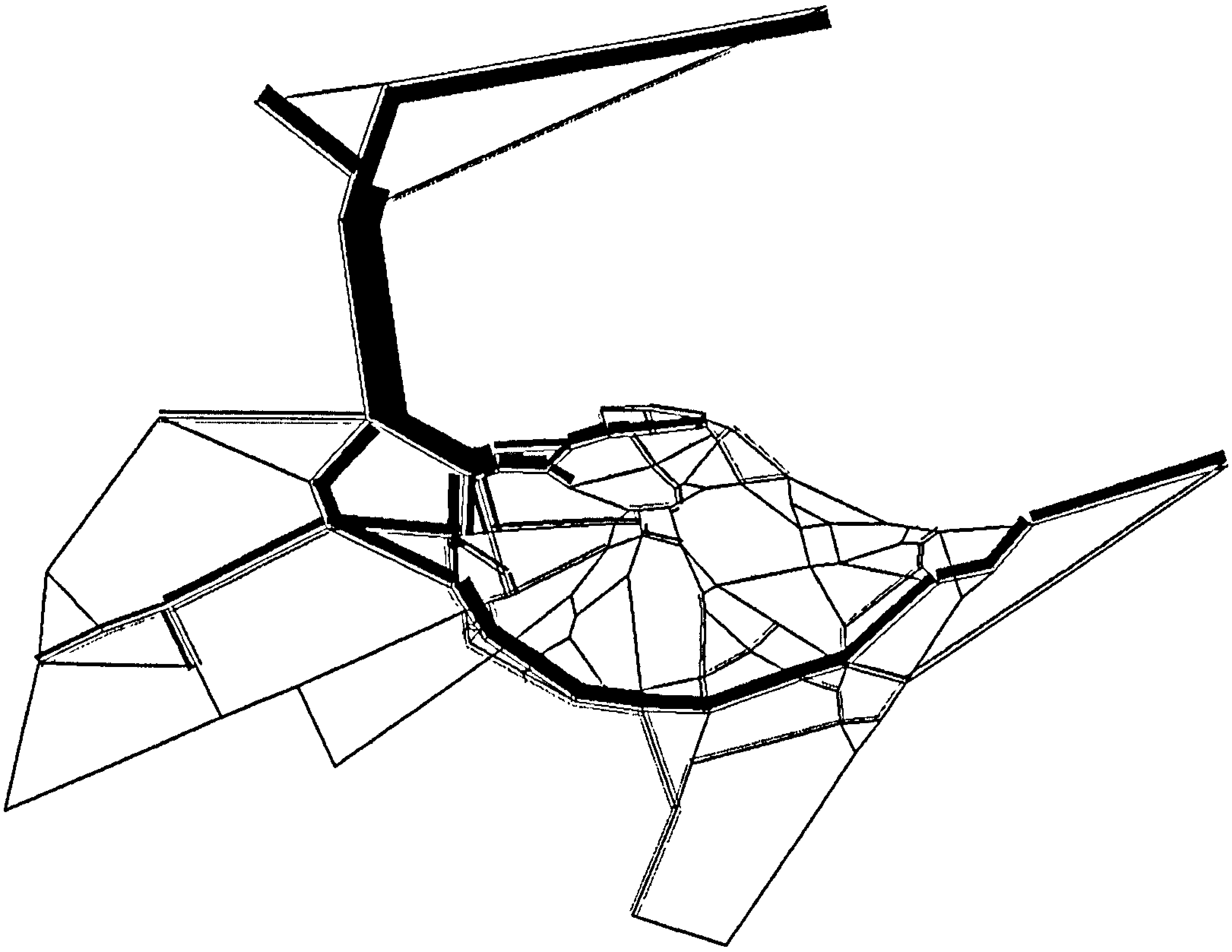


Figure 9-4 Plot of the trips passing through links A and B in the do-nothing scenario

Next, we discuss the effect of the toll location on the performance of the road pricing scheme. From Table 9-2, obviously different locations of the charging cordon produce different levels of benefit. The best judgmental cordon with the uniform toll is the inner2 cordon which generated the net benefit improvement of £3.99k/peakhour. On the other hand, the optimised design of a single cordon inside the city ring road (OPC1) generates the net benefit improvement of £7.21k/peakhour which is about 80% more than the benefit from the inner2 cordon. This clearly demonstrates the significant impact of the cordon location on the performance of the scheme. In addition to the single cordon design, the optimised double cordon scheme (D-OPC) improves the net benefit improvement of the scheme further (with about 112% higher net benefit compared to the OPC1). However, it should be noted that D-OPC imposes the toll on the Forth Bridge which is the link with the highest marginal cost. On the other hand, the OPC1 cordon was allowed to impose the toll only on the area inside the city by-pass which excludes the Forth Bridge.

Figure 9-5 shows the comparisons of the link flows from the tests with the OPC1 and inner2 cordons. The width of the bandwidth represents the difference between the flows in the two cases. As mentioned earlier, the main congestion problem and hence external cost in the Edinburgh network occurs in the west part of the city due to a higher volume of demand compared to other movements in the network. Figure 9-5 demonstrates the different effect of the OPC1 and inner2 cordon in which the OPC1 cordon was located in such a way the trips in the west part of the city centre are affected most (the cordon cover a wider part of the west of the city). Thus, OPC1 depressed more trips in the west part of the city, which is the most congested area, compared to the inner2 cordon. This result implies the advantage of the optimised location of a charging cordon scheme where the optimised location can target better the area affecting the social welfare of the system.

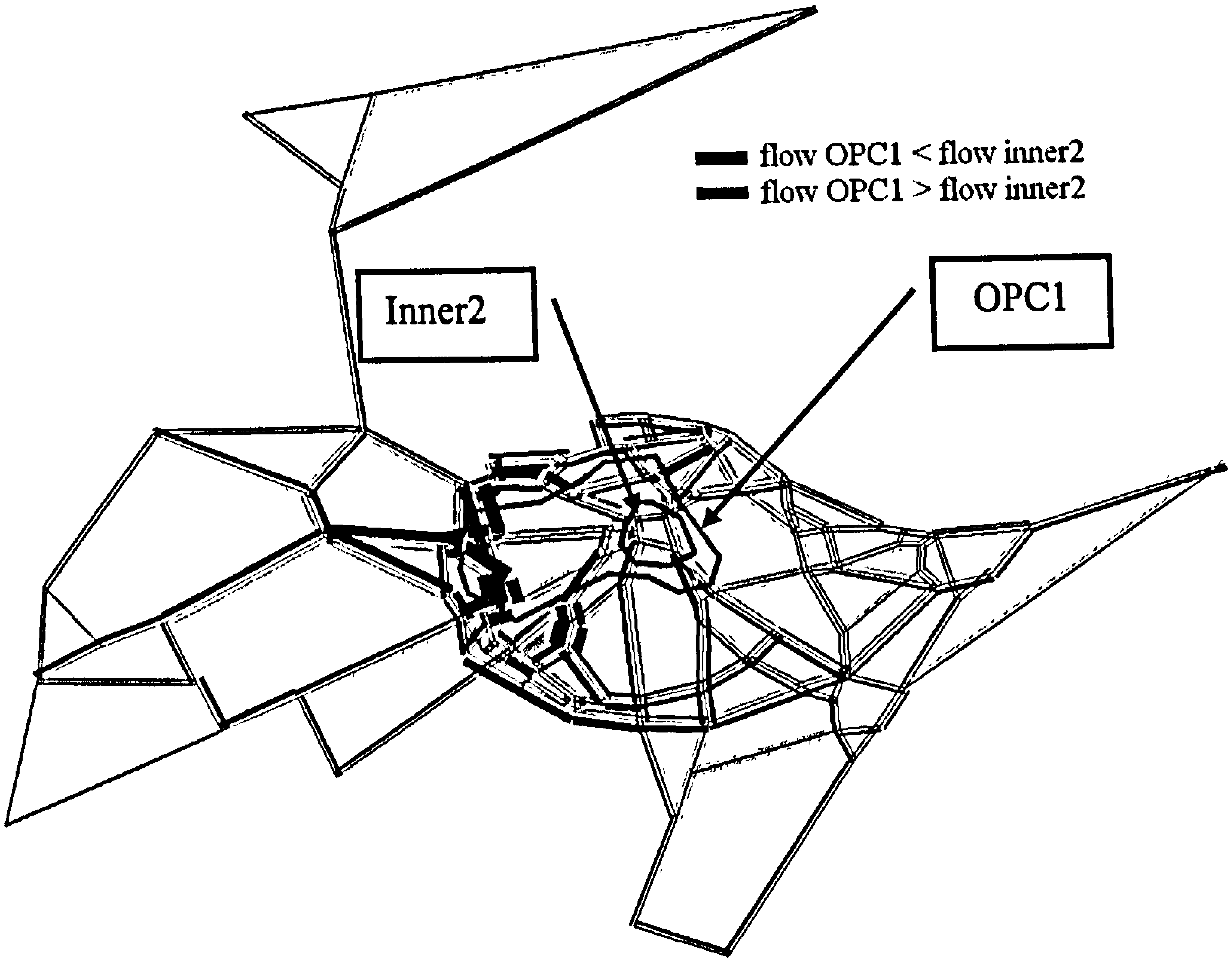


Figure 9-5 Comparison of the link flows from the tests with OPC1 and inner2 cordons

This phenomenon can also be observed by comparing the results between the Top-15 with variable toll and the OPC1 with an optimal uniform toll. With the Top-15 tolled links and variable tolls, the toll scheme generates about 150% higher benefit compared to the benefit from the OPC1 with uniform toll. If we observe the traffic volumes in both cases, the total traffic volume left in the network with the Top-15 scheme is in fact higher than those from the OPC1.

Figure 9-6 compares the link flows from the Top-15 and OPC1 schemes. The figure shows the great reduction of the traffic under the Top-15 toll scheme in some parts of the network, especially on links A and B and the east and north corridors of the city centre. These area and links are highly congested in the do-nothing scenario. On the other hand, the Top-15 toll scheme depresses fewer trips in other parts of the network. This demonstrates the effect of the choice of toll location in which the Top-15 toll scheme (with its variable tolls) can concentrate more on reducing the trips contributing most to the objective function.

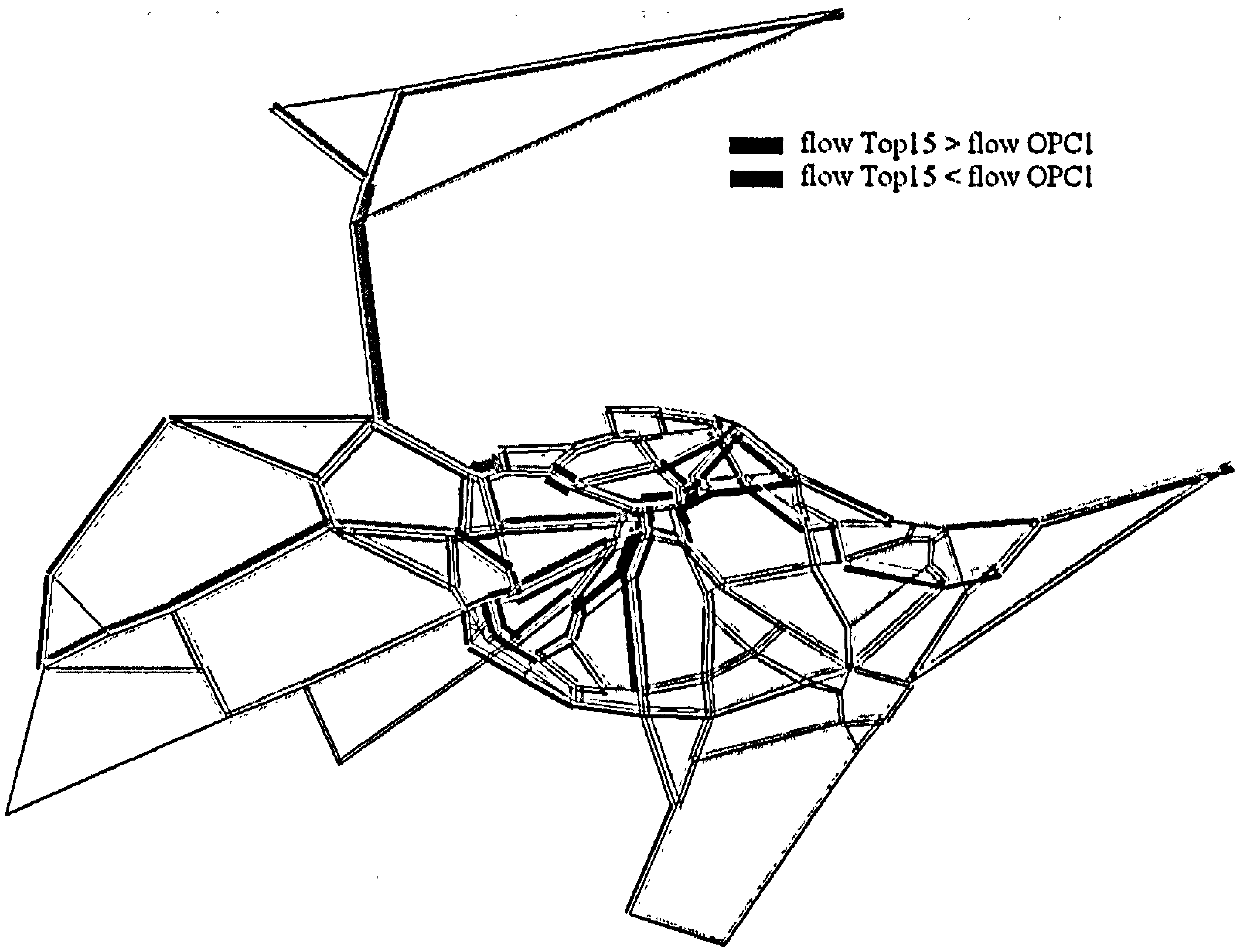


Figure 9-6 Comparison of the link flows from the Top-15 and OPC1 charging scheme

Different cordon designs also generate different adverse impacts. Table 9-2 above shows some important outcomes for different tolling schemes including the demand, the total travel time, total travel distance, and equity impact measured through the *Gini* coefficient. The D-OPC cordon depressed the demand most and also resulted in a low level of total travel distance and total travel time. Interestingly, some form of relationship between the net benefit and the Gini coefficient can be observed in which a scheme with a high level of net benefit seems to have a worse impact on the equity as well (as measured by the level of the *Gini* coefficient). The Top-15 scheme with variable tolls generated the highest net benefit but also has the highest value of the *Gini* coefficient which implies the worst impact on equity (although it does have a higher demand in the network compared to the D-OPC).

Apart from the key statistics presented in Table 9-2, some other interesting outcomes of the cordon design can also be analysed. Table 9-3 presents additional indicators of the effect of different scheme designs including the travel time/trip, travel distance/trip, and average speed in the network.

Charging system	Optimal toll	Travel time/trip (minute)	Travel distance/trip (km)	Avg. speed (km/hr)
Inner cordon 1	£0.50	31.41	16.40	31.33
	varied	31.22	16.42	31.56
Inner cordon 2	£0.75	31.08	16.49	31.83
Outer cordon	£0.75	31.30	16.30	31.25
	varied	31.09	16.32	31.50
OPC1	£1.50	30.30	16.61	32.89
Top-15	£0.75	30.01	16.45	32.90
	varied	27.67	16.19	35.11
D-OPC	£1.25	27.99	16.61	35.61

Table 9-3 Statistics on travel time/trip, travel distance/trip, and average speed for different charging scheme of the Edinburgh network

From Table 9-3, the effect of the cordon location can be observed from the comparison between the outer and the inner2 cordons with the same toll level of £0.75. The inner2 cordon generated the lower travel time/trip but higher travel distance/trip compared to the outer cordon. This result implies that the inner2 cordon which is a narrower cordon affects mainly the short trips on congested parts of the network and generates more diversion (hence increase the travel distance/trip). These two effects can be caused by the size of the inner2 cordon that mainly focuses on the trips entering the city centre and is easy for the trips to divert around unlike the outer cordon that covers a very wide area of the network and trips and is difficult for the trip to divert around the cordon.

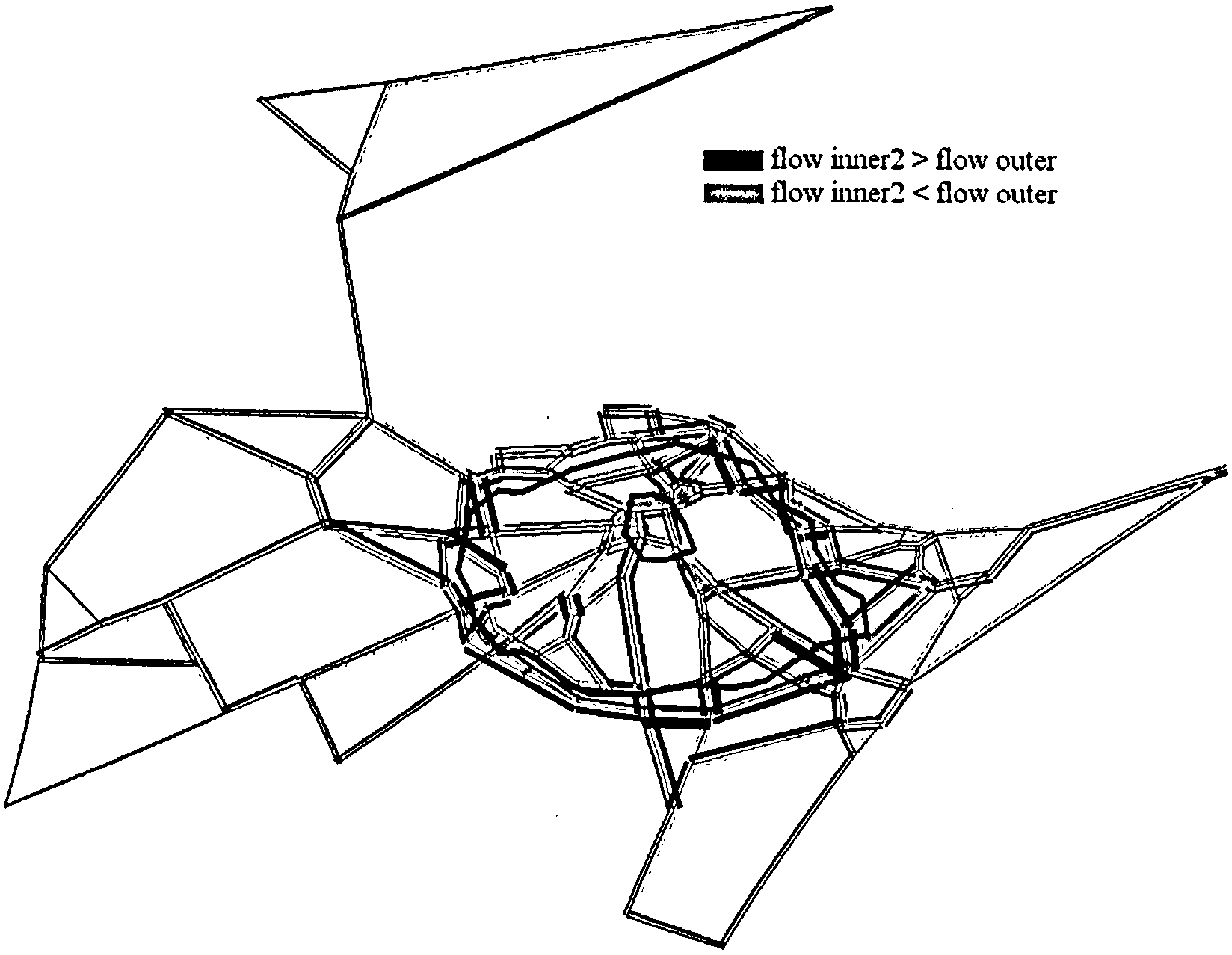


Figure 9-7 Comparison of link flows with the inner2 and outer cordons

Figure 9-7 above shows this effect in which the inner2 cordon generated a higher traffic volume around the city bypass whereas depressed more trips inside the bypass. It could be observed that the main traffic on the bypass with the inner2 cordon was originated from the area inside the bypass travelling to the destination outside the city bypass area. These trips tried to avoid the charges imposed in the city centre area by diverting through the city bypass.

Interestingly, the outer cordon did not generate the same level of the traffic diversion despite the location of the cordon just inside the bypass. One reason could be the substantial reduction of the congestion inside the cordon as a result of which the trips generated inside the cordon could rather travel through this area (they are not tolled by the outer cordon when travelling through the area inside the cordon). For the through traffic from outside the cordon, they are either depressed by the toll level or divert around the charged area.

For a general comparison, the Top-15 scheme with variable tolls generated the lowest travel time per trip and travel distance per trip. On the other hand, the average speed in the network with the D-OPC is the highest which could be due to the lower level of demand in the network. For a single cordon scheme with uniform toll, the OPC1 generated the lowest travel time per trip and highest speed which implies the least congestion level in the network. However, it also generated the highest travel distance per trip.

The effect of the OPC1 on the link flows compared to the outer cordon is very similar to the comparison between the inner2 and the outer cordons described earlier. OPC1 generated a higher level of diversion trips generated from inside the city bypass. At the same time, when comparing the link flow results of the OPC1 and the inner2 cordons it can be observed that the OPC1 allowed more trips initiated inside the bypass to travel through the city centre area (see Figure 9-5 presented earlier). This is due to the larger size of the OPC1 compared to the inner2 cordon.

Overall, the results discussed above stress the importance of the toll level and location. Both toll level and location should be adjusted appropriately to impose the right amount of toll on the parts of the network with or causing high marginal costs. As the results showed, simply locating a charging cordon judgmentally does not guarantee the performance of the scheme. Obviously, a higher flexibility which implies a higher degree of freedom for the toll scheme design will result in a scheme with greater performance. In the next section, some issues related to the trade-off between design constraints and the performance of the charging cordon scheme are discussed.

9.2.2 Trade-off between practical constraints and benefits

Apart from the practical requirement of a charging cordon format, several outcome constraints were also discussed in Chapter 3. The term outcome constraint refers to constraints on the outcome of the policy (e.g. traffic volume, revenue, or equity impact). The recognition of the need to include additional outcome constraints into the design of a charging cordon was made in Chapter 7 in which the CON-GAAS was developed for tackling this problem. The other set of constraints may include the coverage area of the charging cordon (OPC1 vs. D-OPC) and the possible toll level (e.g. some tests with the optimal cordon design with predefined toll level). Note that some of the discussion on this issue was already presented earlier in Chapter 7. The aim of this section is to revisit and stress the trade-off between the benefit of a charging cordon scheme and additional constraints on the design.

Obviously, adding constraints to any design will reduce some of the benefit from the scheme. Table 9-4 summarises the results from the previous chapters of the different optimal charging cordon design problems with different constraints. OPC-£0.50 – OPC£4.00 cordons were those cordon optimised with a predefined uniform toll (e.g. OPC-£0.50 is the optimal cordon location with the toll level of £0.50). These cordons were presented earlier in Chapter 6. CON-REV, CON-GINI, CON-TTTIME, CON-TTDIS, CON-REV-GINI are the optimal cordons with outcome constraints on net revenue, equity impact, total travel time, total travel distance, and both net revenue and equity impact respectively. These cordons were found by CON-GAAS in Chapter 7.

Firstly, the constraint on toll level of the charging cordon scheme is imposed. As discussed previously in Chapter 6, different constraints on the toll level resulted in different levels of scheme benefits. From Table 9-4, apart from the toll level of £1.50 which is the optimal toll of OPC1 the cordon with the predefined uniform toll level of

£1.00 generates the highest net benefit compared to the other cordons with predefined uniform toll. Nevertheless, the net benefit improvement from OPC-£1.00 is around 12% lower than the net benefit achieved by OPC1. As the constrained toll level deviates further away from the optimal toll of £1.50, the level of the net benefit falls further. Ultimately, with the toll level of £4.00, the scheme does not generate any positive net benefit improvement. With this simple result, clearly the traditional way in which the decision maker makes the decision on the toll level prior to the road pricing scheme design may yield a scheme with a substantially lower benefit. On the other side of the coin, it is of practical and political interest to restrict the level of toll. This imposes a serious question on the trade-off between the constraint on toll level and the benefit of the charging scheme. For example, if Edinburgh were to restrict the charge to £1 to increase acceptability, then this would lose 12% of the modelled welfare improvement of optimal toll and hence reduce the effectiveness of the design.

Charging system	**Optimal/ Predefined toll**	**Net EB (£k/hour)**	**Net Revenues (£k/hour)**	**TT. time (PCU-Hr/hr)**	**TT. distance (PCU-Km/hr)**	**Demand level (kPCU/hr)**	**Gini**
OPC1	£1.50	7.21	43.70	52325	1720864	103.6	0.41
OPC-£0.50	£0.50	4.90 (-32%)[*]	18.27	55399	1757031	106.9	0.28
OPC-£0.75	£0.75	5.56 (-23%)	23.79	54799	1746423	106.1	0.36
OPC-£1.00	£1.00	6.35 (-12%)	24.15	53971	1737062	105.1	0.36
OPC-£1.25	£1.25	6.13 (-15%)	37.29	53170	1727763	104.1	0.36
OPC-£2.00	£2.00	5.25 (-27%)	47.67	52030	1720246	103.3	0.45
OPC-£3.00	£3.00	2.88 (-60%)	75.94	49054	1669723	99.6	0.38
OPC-£4.00	£4.00	-1.40 (-119%)	80.04	48679	1679186	100.2	0.52
CON-REV	£2.00	6.99 (-3%)	56.44	51018	1703384	102.2	0.40
CON-GINI	£0.75	5.79 (-20%)	27.16	54328	1740770	105.4	0.28
CON-TTTIME	£3.00	3.75 (-48%)	75.74	48910	1673434	99.99	0.39
CON-TTDIS	£2.00	4.50 (-37%)	56.21	51229	1687498	101.64	0.40
CON-REV-GINI	£1.50	4.38 (-39%)	48.55	52139	1717133	102.5	0.29
D-OPC	£1.25	15.28 (+112%)	101.50	45157	1607555	96.8	0.48

Table 9-4 Comparison of the scheme benefits with additional constraints

---

[*] Percentage increase (+) and decrease (-) in the net EB improvement of the charging scheme compared to the net EB improvement of the OPC1.

On a slightly different perspective, the constraint on the structure of the cordon can also change the benefit of the scheme. As discussed earlier, the optimised double cordon (D-OPC) generates a substantially higher benefit compared to the OPC1. The D-OPC cordon is a double cordon and in the optimisation process the inner and outer cordons were allowed to cover the areas inside and outside the city bypass respectively. On the other hand, the OPC1 cordon is a single cordon and was only allowed to be inside the city bypass.

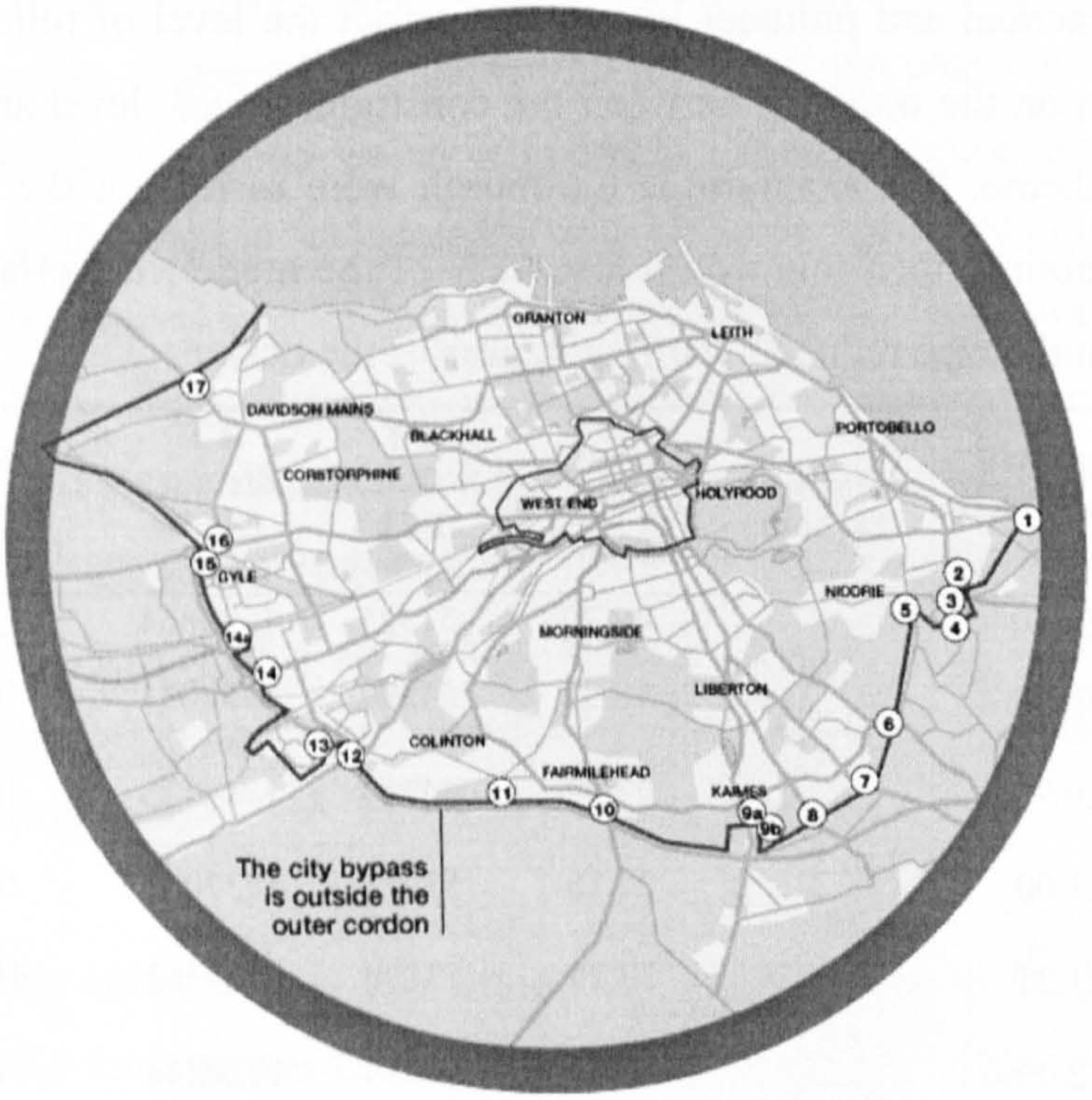


Figure 9-8 Current proposed double charging cordon for the Edinburgh scheme (Source: http://www.tiedinburgh.co.uk/congest.html)

With this result, the decision maker may have to consider the pro and con when deciding on the possible charging area and the structure of the cordon. The current design of the Edinburgh road pricing scheme covers mostly the area inside the city by-pass with double cordon scheme (see Figure 9-8). This is the same area as the OPC1. In this thesis, we did not conduct the test to design a double-cordon scheme inside the city by-pass area due to the lack of the resolution of the network. Nevertheless, it can be shown from the results discussed that the extension of the scheme outside the city bypass may bring a substantial additional benefit to the road pricing scheme. In particular, the key congested link is the Forth Bridge which is one of the tolled links of the D-OPC. However, it should be noted that some of the area outside the city bypass may be outside the authority of the Edinburgh city council. In addition, currently there is already a charge for using the Forth Bridge under the bridge financing scheme, and it is not possible to introduce additional toll. This is indeed a practical constraint discussed earlier in Chapter 3.

Outcome constrains were also tested. Firstly, a constraint on the net revenue was added into the design (net revenue ≥£50k). The initial discussion can be made by observing Figure 9-9 which was also presented earlier in Chapter 6. From the figure, the uniform toll for the OPC1 should be at least £2.00 so that the scheme will be comply with the constraint on the revenue. If the designer simply adapts the OPC1 with this change, the loss in the net benefit improvement will be around 3% of that for OPC1 with £1.50 toll.

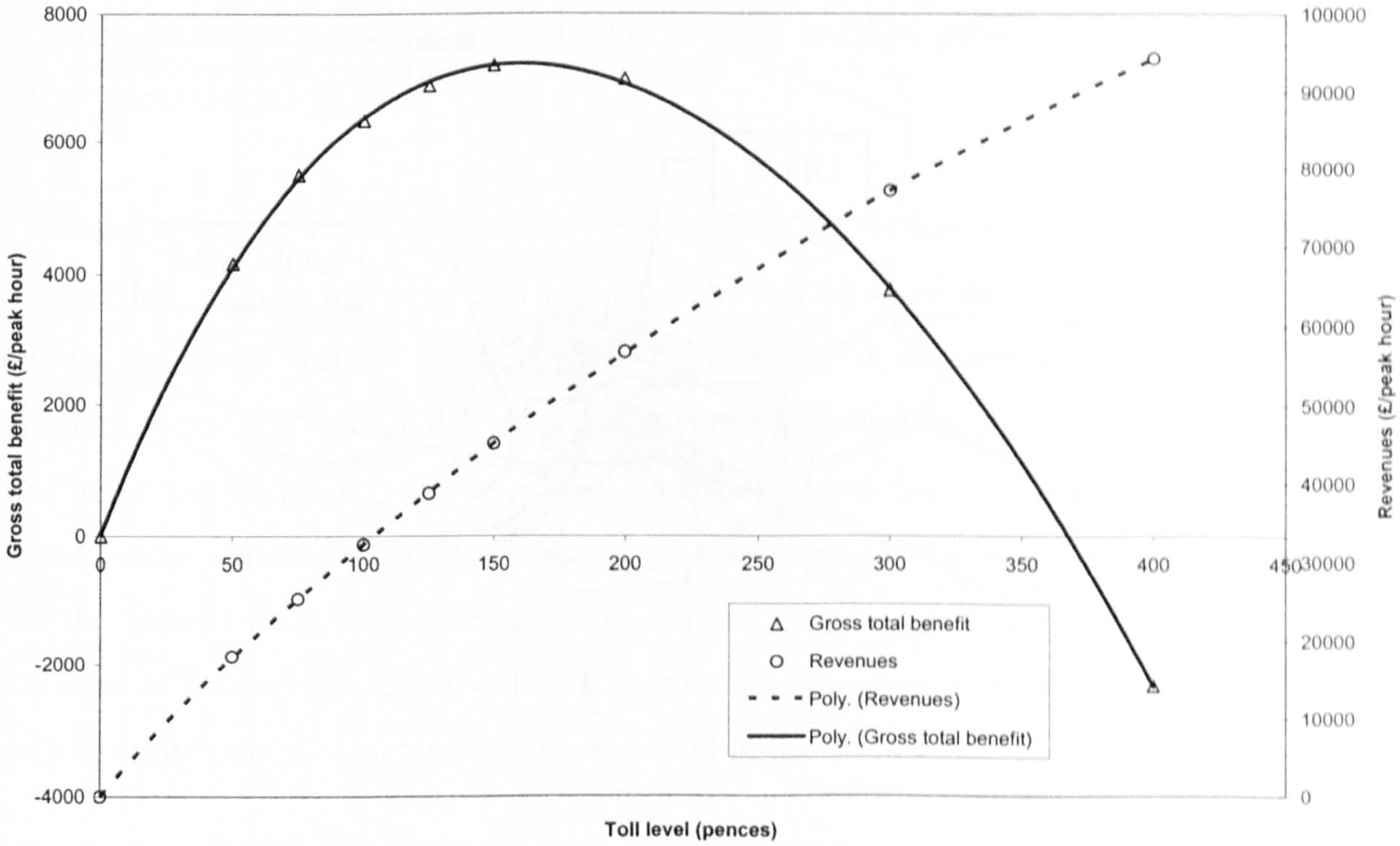


Figure 9-9 Variation of total benefits and revenues for OPC1 cordon with different toll levels

Coincidently, the CON-REV scheme which was found by CON-GAAS is the same cordon, OPC1, but with the toll level of £2.00. This implies that the location of the OPC1 is capable of generating a higher level of revenue and the only adjustment needed is the toll level. This may also be due to the fact that there is a high correlation between the social welfare objective and revenue objective since the revenue is a part of the net benefit formulation (see Equation 4-13 page 94). However, this result may not be applied to other cases in which the modification of the cordon location may be a better solution for the design with revenue constraint.

The second constraint considered was the equity constraint. With the constraint on the Gini coefficient (Gini ≤0.30), CON-GINI (see Figure 9-10) was found by the CON-GAAS as the optimal cordon. The inclusion of the constraint on the equity in this test reduces the net EB improvement (compared to OPC1) by 19%. Again, this result shows the need to trade-off of the benefit of the scheme with additional outcome constraint. Figure 9-10 shows the comparison of the link flows from OPC1 and CON-GINI.

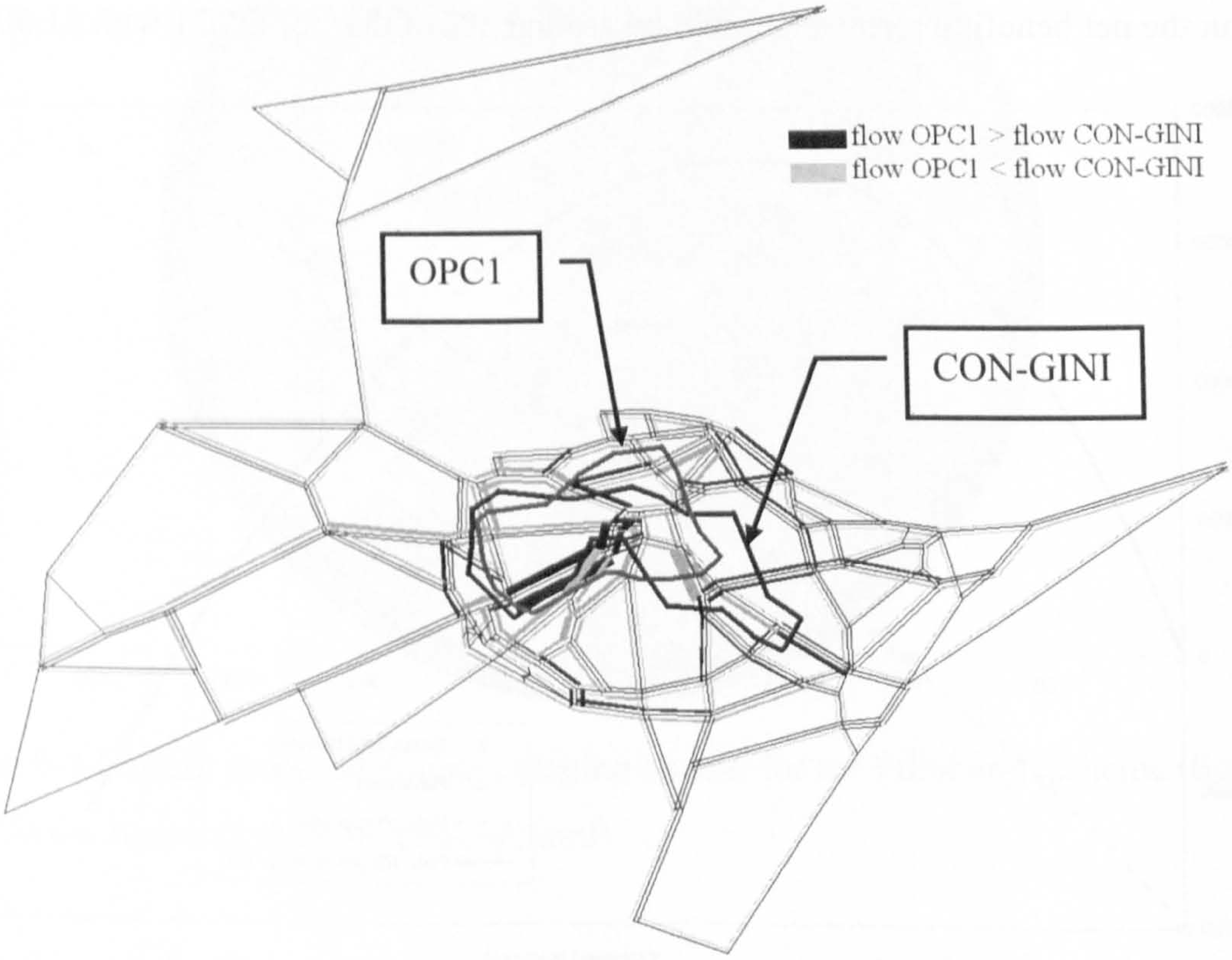


Figure 9-10 Comparison of the link flows from OPC1 and CON-GINI

The concept of *Gini* coefficient is to encourage a scheme with well distributed costs and benefits compared to the do-minimum scenario. The result in Figure 9-10 shows the greater reduction of flows in the case of the CON-GINI in the east part of the city and some of the corridors leading to the city centre. This area was not the main focus of OPC1 due to its low contribution toward the social welfare improvement.

Similarly to the tests with the constraints on the net revenue and equity, the results of the constrained cordon design with the constraints on total travel time, total travel distance, and both net revenue and equity resulted in the reduction of the net benefit (see Table 9-4).

9.2.3 Compromised cordon design with multiple objectives

This section attempts to characterise the features of a charging cordon that are suitable for different objectives. Three main objectives are discussed including the social welfare, revenue generation, and equity. It should be noted that the discussion of equity is limited to the form of equity considered in this thesis (Gini coefficient). The results may not be applicable to other forms or definitions of equity. Some of the discussion was already provided in the previous two sections. In particular, Section 9.2.1 already discussed the characteristic of the toll level and toll location generating a high net EB improvement (social welfare).

In summary, the key characteristic of a welfare charging cordon is to concentrate on the area with the highest marginal cost in which the reduction of the trips will contribute most to the social welfare improvement. The number of trips suppressed is also an important factor; the welfare cordon should depress trips by just the right amount. If too many trips are depressed, drivers whose benefits exceed their marginal costs will be targeted, thus reducing economic benefits of the scheme. This will of course interact with the benefit from travel time saving. For the case of the Edinburgh network, as discussed in Section 9.2.1, the welfare cordon tends to cover a wider area of the west part of the city centre, which is highly congested in the do-minimum scenario.

Although revenue is a part of the net benefit objective, there exist some differences in the characteristic of the cordon that is suitable for the objective of maximising the net revenue. In Chapter 6, GA-AS was used to optimise the net revenue, resulting in the design of OPC-REV (with a toll level of £4 which is the highest level toll allowed in the test). The shape of the OPC-REV, as shown again in Figure 9-11 below, suggests that the revenue cordon should have a high number of crossing points, minimising the chance of the trips to avoid the charge, and impose a very high toll. Obviously, this characteristic of the cordon will not be suitable for welfare maximisation since it will overprice the trips. This is verified by the fact that the OPC-REV generated a very high negative net EB improvement. From the table, there exists the relationship between the toll level and the level of net revenue generated by the scheme. Again, this result is consistent with the result in Figure 9-9 in which a higher toll of the OPC1 generates higher net revenue.

Thus, this result presents an interesting conflict between the characteristic of the optimal cordon for the objectives of net benefit improvement and net revenue. On one hand, a welfare cordon should cover the key congested area of the network aiming to tackling the trips contributing most to the objective function with the appropriate level of toll (not overpricing or underpricing). On the other hand, the revenue cordon should have a high number of crossing points covering a wide area of the network and offering no alternative routes. The toll for the revenue cordon will be a high toll level.

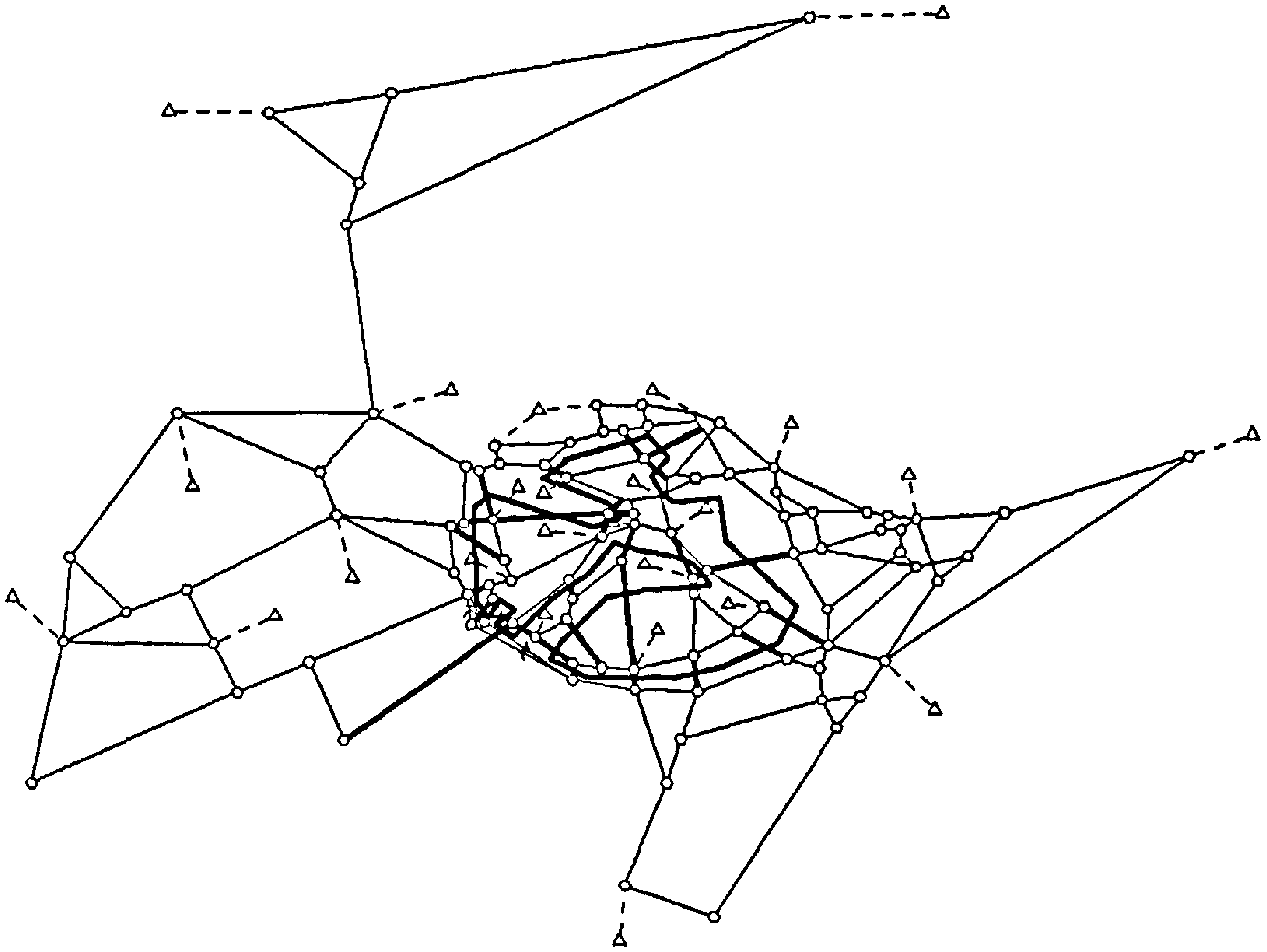


Figure 9-11 Location of OPC-REV as found by GA-AS for maximising the net revenue (highlighted links are tolled links)

In a real world scheme, the revenue of the scheme is one of the key design criteria to ensure the self-financed road pricing scheme and sufficient fund for improving other transport infrastructures. The scheme in London (even with £5.00 toll) generated lower revenue than forecasted which caused some debate on the design of the scheme. In fact, the London ALS, whose design was mainly influenced by political consideration, may be too small. The compromise between the two objectives can be exemplified by the result of the CON-REV cordon which is the same cordon location as the OPC1 but with a higher toll level. Similarly, the design of the compromise cordon for the real world scheme may be enhanced by the application of the algorithms proposed (CON-GAAS and GAAS with NSGA-II) in this thesis.

The last objective is equity as measured by the *Gini* coefficient. As discussed briefly in the previous section, the equity cordon should try to distribute the benefit of the charging scheme equally to all traffic movement in the network. Some results that can be used for the discussion on the main characteristics of the equity cordon are those from the MOOP with the objective of maximising the net benefit and minimising the *Gini* coefficient, as presented in Table 8-3 (page 274). One trend that can be observed from the result is the optimal toll level. The social welfare cordon tends to have a moderate level of toll (around £1.50). This toll level is in fact mainly appropriate for the targeted trips contributing most to the social welfare improvement which are mainly the trips in the congested areas. However, it may be too high for the trips from the less congested areas.

The WG1 and WG31 from Table 8-3 (page 274) can be used to explain this situation (note that WG31 is the OPC1 cordon). Figure 9-12 shows the location of the WG1 cordon which is the cordon with the lowest *Gini* coefficient (hence lowest equity impact) from the set of all Pareto solutions. The toll level for the WG1 is £0.50. Figure 9-13 and Figure 9-14 shows the benefits for each O-D pair under the WG1 and OPC1 cordon respectively. As shown in both figures, the OPC1 cordon generates a high benefit for some of the O-D pairs whereas the O-D benefits in the WG1 case are lower and more similar to one another. In addition, there are also less O-D pairs with disbenefits from the cordon.

This result explains the point about the characteristic of the welfare cordon and the equity cordon discussed earlier. With the OPC1, the cordon focuses on generating the benefit for some particular O-D pairs (those O-D pairs that will contribute most to the social welfare improvement). This will naturally create an equity problem due to its unequal treatment. On the other hand, the equity cordon, WG1, does not generate a substantial a higher benefit for any particular O-D, hence reducing the unequal treatment. The equity cordon tends to cover a wider area of the network to distribute the cost and benefit of the scheme to different O-D pairs, see for example the WG1 cordon compared to OPC1. Given the asymmetric condition of the congestion and travel demand in the Edinburgh network, the equity cordon that covers a wide area of the network will need to adopt a low level of toll to smooth the cost and benefit for different traffic movements.

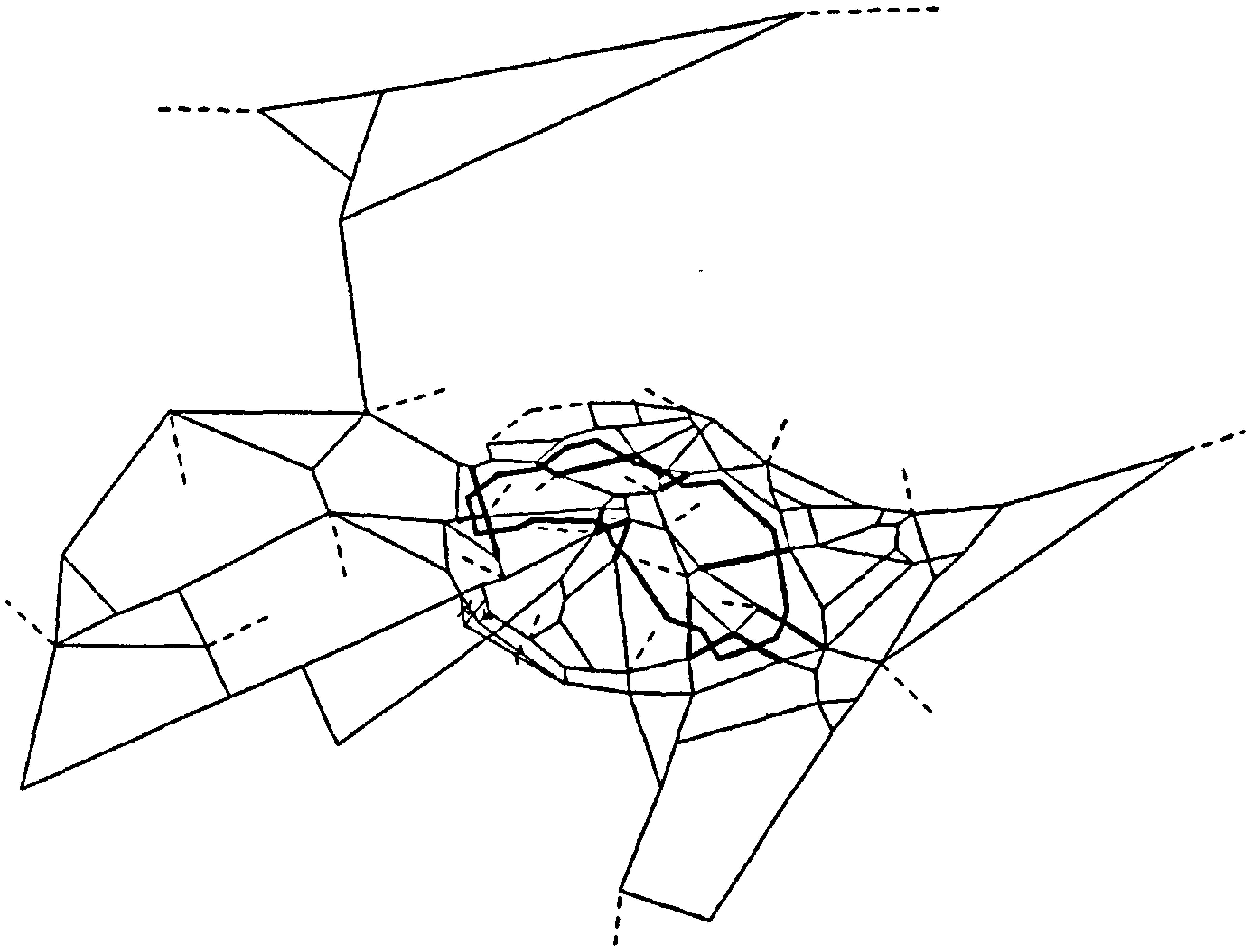


Figure 9-12 Location of the WG1 cordon (highlighted links are tolled links)

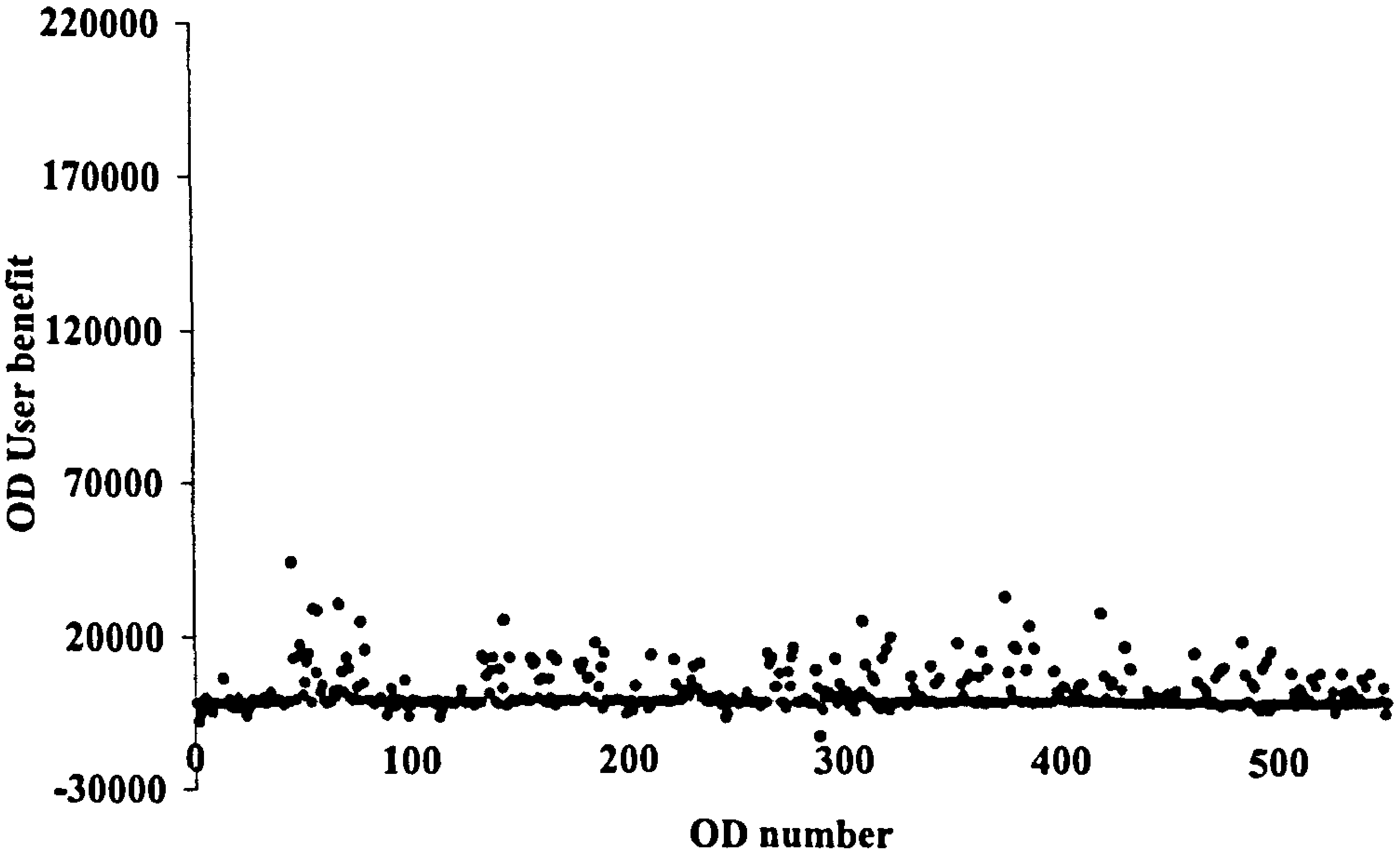


Figure 9-13 O-D benefits with the WG1 cordon (£0.50 toll)

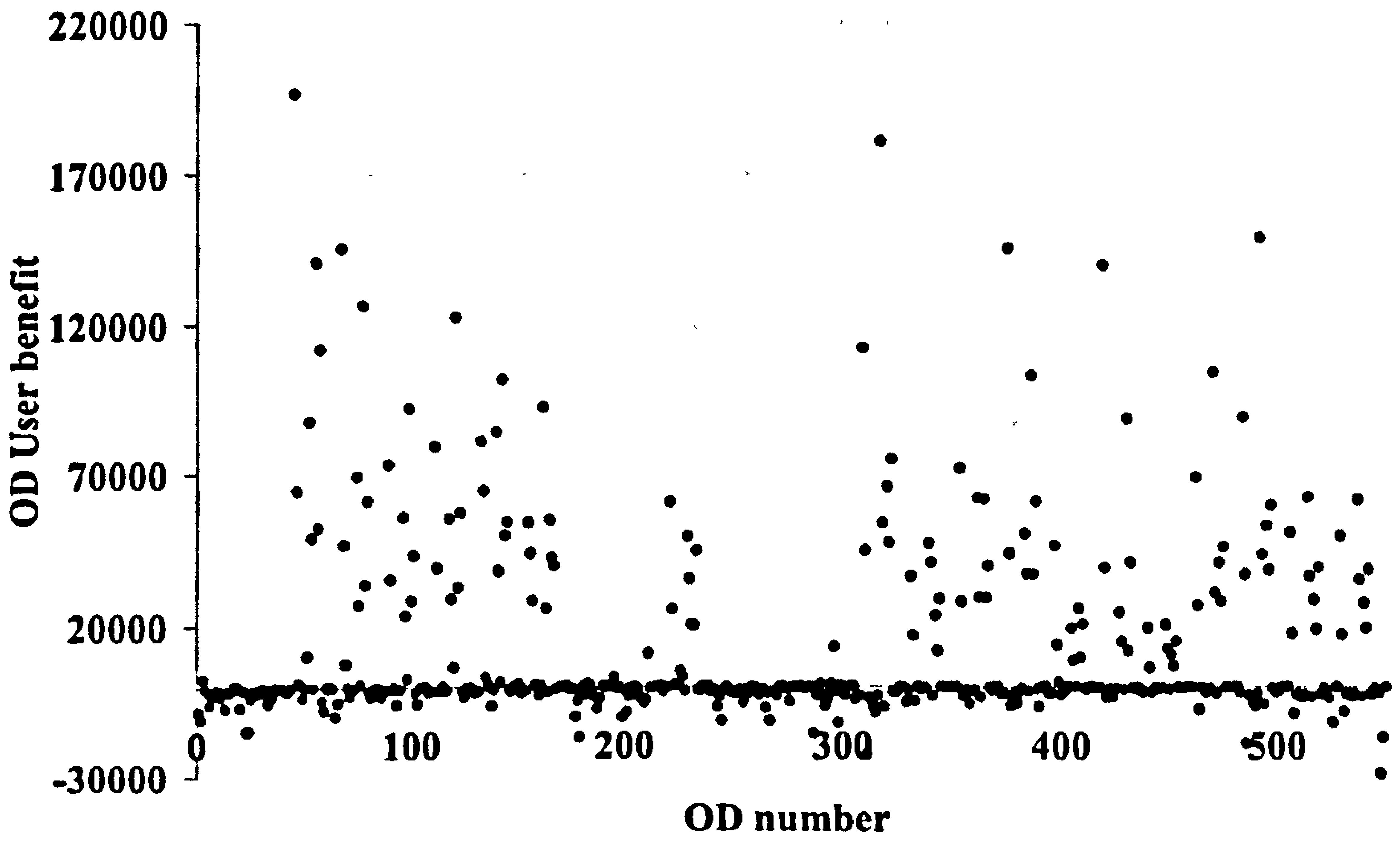


Figure 9-14 O-D benefits with the OPC1 cordon (£1.50 toll)

Overall, the results discussed show clearly the different requirements of the cordon design for different objectives. Similar to the case of the constrained design, the role of the decision maker is to make the decision upon the trade-off between the achievements for different objectives in the design.

9.3 POSSIBLE APPLICATIONS AND BENEFITS OF THE APPROACH TO REAL WORLD DECISIONS

The algorithms developed in this thesis may be able to support the decision making process for the local authority and other responsible bodies to design and implementation road pricing schemes. A number of possible applications and benefits of the methods proposed in this thesis can be suggested as follows.

Providing theoretical benchmark for economic evaluation

The idea of road pricing was originated from the concept of economic efficiency. Thus, the main benefit of the introduction of a road pricing scheme is the increase in the efficiency of the transport system. The question can be asked regarding the magnitude of the benefit road pricing brings into the transport system. From the pure theoretical interest, the concept of first-best pricing policy (all links can be tolled) provides the first bench-mark of the level of the benefit from the introduction of road pricing. However, as discussed in Chapter 2 the concept of first-best pricing is rarely found in the real world and indeed the transaction and technological cost of the road pricing system will reduce the overall benefit of the scheme. For instance, with the Edinburgh network presented earlier, the system optimum marginal cost tolls produce a gross benefit of around £37.8k/peak hour using 320 tolled links. After deduction the cost of scheme implementation and operation (as assumed earlier in Chapter 5), the net benefit improvement from this first-best scheme is around £5.78k/ peak hour. Obviously, this is not a good benchmark for the net benefit of the road pricing scheme anymore.

The second-best pricing strategy, in which only a subset of links in the network will be tolled, was developed to give a more realistic benchmark of the benefit of road pricing taking account of the cost of the scheme. However, as discussed in Chapter 2 and Chapter 3, a more practical road pricing scheme may not even be the second-best toll system. A number of practical requirements need to be taken into account. One possible practical scheme is a charging cordon scheme. This pricing scheme can be considered as the 'third-best' pricing system.

The GA-AS method developed in Chapter 6 can identify the third-best pricing scheme without outcome constraints. GA-AS can be used to seek the optimal charging cordon scheme with a uniform toll producing the benchmark of the benefit achievable by a cordon pricing scheme. This benchmark can then be used to identify the loss of benefit from the 'third-best' road pricing scheme with additional design criterion.

Formalise the design process of a road pricing scheme

There are various scheme designs of road pricing cordons around the world as discussed in Chapter 2. However, the actual process and reasoning of the design of these schemes is not well documented. As analysed in Chapter 3, this process relies on the professional judgment and experience of the designer. The key drawback of this approach discussed throughout this thesis is the possible loss of the benefit from the charging cordon scheme by choosing a sub-optimal scheme. In addition, a set of outcome constraints cannot be included into the design process easily apart from in the post-evaluation process. The post-evaluation process simply checks the impact of the predefined charging system from the output of an appropriate traffic model. If none of the predefined set of designs can satisfy the set of outcome constraints, then some modification to the design will be made, again, judgmentally. The traditional method for designing a road pricing scheme can thus be described as a trial and error method.

The mathematical and computerised approach proposed in this thesis allows a more direct and formal treatment of the objectives of the scheme and the constraints. As shown in Chapter 6, the cordon design (without constraints) from the GA-AS method outperforms those from the judgmental design. In addition, any outcome constraints of the design can be included into the consideration explicitly during the optimisation process as shown in Chapter 7 (CON-GAAS).

The advantage of the algorithm developed is not only the optimality of the design and its ability to deal with constraints, but also the formalism and transferability of the method of this approach. The formalism means that there is well defined process for designing the scheme and a clear reason supporting or justifying the final outcome design (quantitatively). In the case that some qualitative criteria are involved in the design, these criteria must be defined explicitly in the format that the optimisation routine can consider (e.g. the mathematical form of closed cordon requirement). This will increase the transparency of the design process of a charging cordon scheme in the sense that all criteria involved in the design must be defined explicitly.

In addition to the formalism of the process, the knowledge on how to design a road pricing scheme with a well defined tool (like GA-AS and CON-GAAS) can be transferred easily. In the traditional approach, one may need to spend a substantial amount of time before becoming an experienced designer of road pricing schemes. This is of course linked to the lack of the formalism of the judgmental approach. On the other hand, with a clear procedure and tool to support the design, the knowledge on how to design a charging cordon scheme can be transferred from one group to others and from one case to other cases easily.

Exploratory design with incremental constraints and criteria

As mentioned, the algorithm developed in this thesis allows the explicit inclusion of various constraints into the design process. Although the scheme designer and decision maker may have a clear idea of the list of possible design constraints and criteria, some unknown adverse impacts of the scheme may be revealed later on from the modelling output. In the traditional or judgmental process, it is rather a difficult task to introduce a new set of constraints/criteria into a design and come up with new design options. On the other hand, the automatic computerised design process proposed in this thesis can be recalled with additional constraints to generate a new set of possible cordon designs. This ability allows the decision maker and the scheme designer to explore the possible adverse impacts of the scheme which may not be obvious and then incrementally include them into the design. Similarly, the algorithm can also be used to explore the effect of different constraints on the design in order to study and explain to the decision maker thoroughly the trade-off of the benefits of the scheme and the constraints added into the design.

For example, the current issue of the unfair treatment of residents outside the Edinburgh city region, arising from the double cordon design, was raised after the city proposed the scheme to the public. However, as mentioned it is difficult to include this kind of outcome constraint into the actual design process of the charging cordon scheme. With the CON-GAAS and MOOP-GAAS, the designer of the Edinburgh scheme may be able to explore the effect of the inclusion of this new constraint/criterion as soon as it is raised and input the design result into the discussion.

Flexible trade-off process between the different criteria

From the review in Chapter 2 and interview with the local authorities in Chapter 3, it is clear that road pricing is an urban management tool serving several objectives. The idea of designing a scheme that is suitable just for one objective may not be practical. Thus, this requires the consideration of multiple objectives (or multi-criteria) in designing the scheme. The traditional approach for dealing with multi-criteria in transport planning has been attached mainly to the appraisal technique (e.g. using the preference technique to assign the weights to different criteria). However, as discussed extensively in Chapter

8 this is not the only available approach for aiding the decision making process. Indeed, several advantages could be gained by adopting a rather non-conventional approach. The MOOP-GAAS was proposed in Chapter 8 for dealing with the multiobjective design of a road pricing scheme using a multiobjective optimisation approach.

The MOOP-GAAS developed allows a more flexible trade-off process between different objectives of the scheme. The decision maker is equipped with real trade-off information before actually making the decision on the trade-off between different objectives (through the set of the Pareto solutions generated by the MOOP-GAAS). Also, with a clear output in the form of the trade-off surface the decision maker (and the city) can communicate with the involved parties easily on the issue of the conflict between the requirements from different groups. For instance, the discussion for the current extension plan of the London congestion charging scheme can be aided by the information on the actual trade-off between the amount of revenue generated from different design and the welfare improvement.

9.4 SUMMARY

This chapter digests the numerical results from the previous chapters to investigate the effect of the toll level and location on the performance of the scheme, the trade-off between the design constraint and the scheme performance, and the multiobjective design.

The analysis suggests, as expected, the significant influence of the toll level and location of the charging cordon scheme on its contribution to net benefit improvement. The asymmetric nature of the traffic condition (and hence marginal cost) in the traffic network requires some degree of differentiation of the toll imposed on different trip movements in order to reflect their marginal costs (which is consistent with the concept of a social welfare maximisation toll). The variable toll level scheme provides a higher degree of freedom for adjusting the toll levels for different groups of traffic. From the tests, with the same toll location the scheme with optimised variable toll (by GA-CHARGE) can generate up to 95% higher benefit compared to its counterpart uniform toll scheme (this is the case of the Top-15 tolled links).

The careful selection of the toll location can bring additional benefit. A well designed scheme, like OPC1, with a uniform toll in our example can generate around 80% higher benefit compared to the best judgmental cordon (inner2 cordon). The fact that the marginal cost and congestion level varies across different parts of the network also implies that the pricing scheme focusing on the key areas (contributing most to the net benefit improvement) can increase the benefit of the scheme significantly.

The analysis of the previous results in this chapter also confirms the loss of the benefit due to additional design constraints. This is correlated to the fact that there exist the differences in the characteristics of the charging cordon that are suitable for different objectives. The welfare cordon should concentrate on the groups of traffic with the highest marginal cost (implying the coverage of the scheme on the congested part of the network), and the toll level should be at the appropriate level for those groups. On the other hand, the revenue cordon should try to impose a high number of crossing points covering a wide area of the network and minimising diversion routes. The cordon should also be associated with a high level of toll to maximise the marginal revenue collected per trip.

The characteristic of the equity cordon is much more complex. The underlying purpose of the *Gini* coefficient (used to measure the equity impact in this thesis) is to distribute the cost and benefit of the scheme in the form of O-D benefit equally to all trips from different O-D pairs. This requirement is clearly in conflict with the welfare cordon described earlier. The comparison made between the OPC1 and WG1 cordon shows this conflict in which the OPC1 cordon creates high benefits for only some of the O-D pairs. Again with the asymmetric nature of the traffic condition in the Edinburgh network, the equity cordon tends to cover a wide area of the network and impose a low toll level.

The chapter also discussed the different potential applications of the methods developed in this thesis. The GA-AS is believed to be able to produce a good benchmark for the benefit of the charging cordon scheme. In relation to this point, the CON-GAAS can then be used to analyse the possible trade-off of the constraint and benefit of the scheme. Similarly, the MOOP-GAAS can be used to aid the decision making over different design objectives. All of these computerised methods impose a formalism onto the design process of the charging cordon scheme in contrast to the informal judgmental design process. The formalism of the process should ease the explanation and justification of the final design afterwards. In addition, the algorithm developed allows the direct inclusion of outcome constraints into the design unlike the judgmental approach. This feature permits the exploratory design of the charging cordon scheme in which additional constraints discovered later on during the design process can be included into the analysis directly.

Overall, the results from the tests of the algorithms developed in this thesis provide interesting information on the charging cordon design. The methods could potentially be used to aid the decision and design process of a road pricing scheme in the real case. However, a number of modelling assumptions were made to simplify the problem at this stage and also additional design features of the charging scheme may be required (e.g. exemption, screenline design, revenue recycling, and combination with other policies). These are envisaged to be the future direction of this research to increase the applicability and realism of the methods. These issues will be discussed in the next chapter. The next chapter will also conclude the thesis by making a number of comments on the development of the algorithm, policy implications of the research, and other relevant possible research topics.

CHAPTER 10 CONCLUSIONS AND FURTHER RESEARCH

The end of one book is the beginning of another.

10.1 THE GAP BETWEEN PRACTICAL AND OPTIMAL ROAD PRICING DESIGN

Recently, there have been a number of key developments on the implementation of road pricing in the U.K. including the change of legislation to allow the introduction of road user charging in 1999, the establishment of the congestion charging development partnership by the UK Department for Transport to stimulate the development of schemes, the commencement of the toll scheme in Durham in 2002 and the London congestion charging scheme in 2003 and, and the initial discussions of the implementation of various schemes in Edinburgh and Bristol. Thus, there is a clear need for a plausible analytical tool that can be used to aid the decision maker in designing the most appropriate road pricing scheme in his/her city.

From the literature, it was found that research to develop optimal road pricing schemes has concentrated on the theoretical construction of the 'second-best' toll pricing system. On the other hand, there exist a number of political and practical requirements for a road pricing scheme found in the literature and from the survey conducted in this study. With these complex practical and political requirements, the design of a road pricing scheme has mainly been based on the professional judgment or experience of the designer. Traffic models have been used as post-evaluation tools for the predefined scheme options.

The theoretical approach has concentrated mainly on the optimality of the road pricing design and as a result has lacked the consideration of practical requirements. On the other hand, the judgmental approach has paid more attention to the practicality and acceptability of the scheme with the risk that it may reduce the benefit of the road pricing scheme. This gap between the research and practical development of the design of a road pricing scheme has been the motivation of this research.

10.2 JUDGMENTAL DESIGN APPROACH

In the first phase of the study, two parallel research strands were carried out. The first is the investigation of the practical requirements of a road pricing scheme. The second is the study of the formulation and solution algorithm for the optimal road pricing design from a theoretical perspective.

From the review, the most commonly used format of road pricing is the charging cordon based scheme in which the users will be charged per crossing of the cordon boundary. This format was believed to be the simplest and most user-friendly. In addition, there exist a number of available technologies for the real world implementation of such schemes. The charging cordon schemes currently under operation include the Singapore road pricing scheme, the toll ring schemes in Norway, and the London Congestion Charging scheme. There are also various plans for the implementation of a cordon based charging scheme currently under discussion in Edinburgh and Bristol.

Given the limited documentation of the judgemental approach, a series of questionnaire surveys and interviews were conducted with a number of cities who were involved in the DfT congestion charging partnership. Most of the cities believed a simple charging cordon scheme to be a good starting point of the road pricing policy with the possibility of modification and extension of the scheme to gain more benefit later on. The cities clearly concentrate much more on the issue of public acceptance and adverse impacts of the scheme than on economic efficiency.

Three issues involved in the judgmental design process were found by the survey including (i) the existence of constraints on the topology and structure for a charging cordon scheme (to avoid public opposition and increase the practicality), (ii) additional consideration of the adverse impact of the scheme design (e.g. equity impacts), and (iii) consideration of a number of objectives rather than an overall efficiency objective. Obviously, with these complex constraints and criteria involved in the design of a road pricing scheme it is difficult for the city to pay much attention to the performance of the scheme. It was found that most cities rather aim to find just a "good enough" cordon to start up the scheme.

10.3 THEORETICAL DESIGN APPROACH

The second and major strand of the study is related to the formulation and solution algorithm of the optimal road pricing design problem. The behaviour of the travellers (and potential travellers) in the network are assumed to follow the well-known concept of Wardrop's user equilibrium (UE). The optimal toll design problem can then be framed as a Mathematical Program with Equilibrium Constraint (MPEC) in which the equilibrium constraint of this problem is the UE condition. The design variables of the optimisation problem may include the toll level and/or toll location. MPEC is a very challenging optimisation problem and still a very active research area. It can also be considered as a Stackelberg game.

Three main objectives were formulated including social welfare, revenue, and equity. The social welfare improvement is defined following the Marshallian measure which is equal to the consumer surplus minus user cost. Similarly, the social welfare improvement can also be calculated by using the rule of half (equal to user benefit plus revenue). With the model assumption of a single class user, only the horizontal equity impact was considered in the study. The equity impact was measured by the distribution of cost and benefit from the road pricing scheme over the trips from different O-D movements. The formulation of the equity impact follows the *Gini* coefficient which is commonly used in the study of income distribution. It was found that the issue of revenue recycling is a very complex issue. Thus, the benefit from the revenue recycling was discarded from the calculation of the equity impact in this study.

The main problems with the MPEC are the non-differentiability of the nonlinear constraints resulting from the equilibrium condition and the linear dependence condition of the active constraints at all feasible solutions. These two problems pose a serious obstacle to the application of nonlinear optimisation algorithm to the MPEC. Different strategies for solving the MPEC rely on different tactics for dealing with the UE condition. The UE condition can be reformulated as an equivalent optimisation problem, mixed complementarity problem, variational inequality, or gap function. With the equivalent optimisation problem, the MPEC becomes a bilevel optimisation programming problem (BLPP). The development of solution algorithms for the optimal toll problem (the design variable is the toll level given the toll location) in the transport context has been a major research area. Several approaches have been proposed including (i) iterative optimisation approach, (ii) linearization approach, (iii) sensitivity analysis, (iv) inverse optimisation, (v) minmax/maxmin based optimisation, (vi) smoothing approach and reduced UE condition, (vii) cutting plane algorithm, and (viii) meta-heuristic optimisation approach.

Most of the algorithms proposed in the transport area stem from the development of the algorithm for MPEC in the area of nonlinear optimisation. From the review, there are four main groups of the optimisation strategies for the continuous MPEC including (i) disjunctive nonlinear programming based algorithm, (ii) penalty and interior point based method, (iii) implicit programming based method, and (iv) nonlinear optimisation based method.

Three possible algorithms were proposed to tackle the optimal toll design problem including (i) the merit function based approach, (ii) the improved cutting plane algorithm, and (iii) the Genetic Algorithm based algorithm. It was found that even with a theoretically well constructed algorithm like the merit function based approach there exists a practical hurdle with the optimal toll design problem concerning the dimension of the problems, which is due to the structure of the network. For a reasonable size network, the number of variables increases rapidly with the number of O-D pairs and links. This is indeed a drawback of solution algorithms applying existing nonlinear optimisation algorithms directly to the reformulated optimal toll problem.

In order to develop a more practical, sound algorithm, the solution algorithm needs to exploit the structure of the problem further. The original cutting plane algorithm proposed for the optimal toll problem indeed exploits the structure of the network and defines the equilibrium condition as a variational inequality with the feasible flow region represented by a set of extreme points. The algorithm was advanced further by introducing the cuts in the flow conservation constraint in the upper level. This modification results in a significant reduction in the number of variables involved in the optimisation problem. However, the improved cutting plane algorithm may not guarantee the optimality of the final solution and should be considered only as a heuristic method. The results from the test of the algorithm with the Pacman network are very promising.

The third algorithm was based on a kind of meta-heuristic optimisation algorithm, Genetic Algorithm (GA). GA does not rely on any properties of the functions involved in the problem. Thus, the GA can be applied directly to the optimal toll problem. The solution of the problem is defined as a chromosome representing different toll levels. Then, for each chromosome a traffic assignment software is used to evaluate the benefit of the tolls. The associated benefit is then assigned as the fitness of that chromosome. The chromosomes will then go through the selection, crossover, and mutation process. The GA based algorithm developed for the optimal toll problem is named GA-CHARGE. GA-CHARGE was tested with the Pacman network and proved to be able to identify similar solutions as found by the improved cutting plane algorithm. In addition, GA-CHARGE was also tested successfully with the Edinburgh network. Despite the successful application of GA-CHARGE to the optimal toll problem, the algorithm was found to be very time consuming. A more deterministic search algorithm like the improved cutting plane algorithm may be considered more efficient for solving the optimal toll problem for a given set of tolled links.

The results from the test of GA-CHARGE with the Edinburgh network also reveal a key comparison between the judgmental design approach and the theoretical approach. The toll levels for two judgmental cordons were re-optimised by the GA-CHARGE and the results suggest that the benefit of the charging cordon with the optimised variable tolls around the cordon can be up to 75% higher than the cordon with an optimal uniform toll. This result implies the greater benefit of the scheme from the theoretical design as expected.

10.4 INTEGRATION OF THE JUDGMENTAL DESIGN AND THEORETICAL DESIGN

After the exploration of both the judgmental and theoretical approaches, the study then attempted to bring in the ability of the theoretical design to enhance the practical design of road pricing scheme. Although the deterministic based methods could be more efficient, the GA based algorithm allows the inclusion of various arbitrary forms of constraints. This is indeed an important requirement for the development of algorithms. Firstly, the GA based algorithm was developed to deal with the design of a road pricing scheme with topological constraint. The most complex topological constraint dealt with in the algorithm is the requirement of the cordon format. The concept of a branch-tree framework was devised to encapsulate a charging cordon in two strings. These two strings are then used as a chromosome in GA.

The algorithm developed for tackling the optimal charging cordon design was named GA-AS. GA-AS is capable of designing an optimal closed charging cordon scheme (applicable also to multiple cordons) with an optimal uniform toll. The structure of the algorithm was similar to GA-CHARGE in which a traffic assignment software is used to evaluate the fitness of the cordon.

GA-AS was tested successfully with the Edinburgh network. The resulting design (OPC1) outperformed the best judgmental cordon producing around 80% higher benefit. GA-AS was also adapted to optimise the design with a different objective to maximise the net revenue. This was mainly to demonstrate the flexibility of the algorithm for dealing with different objective functions. The algorithm was also tested with different parameters for the GA including the probability of mutation (Pm) and probability of crossover (Pc). It was found that the performance of GA depends significantly on the setting of Pm and Pc. This is the main drawback of a meta-heuristic search technique like GA in which the user may need to define many parameters before running the algorithm. If one adopts an inappropriate setup of these parameters, the performance of GA may not be optimised.

The development of GA-AS represents the first step toward the integration of the theoretical and judgmental approaches for the design of a road pricing scheme. The next development was to allow for the inclusion of outcome constraints on the scheme design. The GA-AS algorithm was extended to deal with several outcome constraints. Four main constraints were tested including the net revenue, equity impact, total travel time, and total travel distance. From the review, three categories of methods for extending the GA-AS to deal with the constraints were available including the constraint relaxation approach, the interior feasible space search, and the hybrid method.

The interior feasible space search is mainly suitable for constraints related to the structure of the solution, e.g. the closed cordon formulation requirement. However, this approach may not be appropriate for constraints related to the outcome of the solution since it is difficult to construct a searching mechanism that can ensure the feasibility of the solutions at all times. The hybrid method was considered too complex but may be applicable to the problem. An approach based on the constraint relaxation method was ultimately chosen for dealing with the constraints. In particular the self-adaptive penalty based algorithm was implemented with GA-AS (the method is named CON-GAAS).

CON-GAAS was also tested with different problems again with the network of Edinburgh. In all tests, there were different levels of reduction in the net benefit compared to the net benefit from the optimal cordon without any constraint (OPC1). The constraint affecting the net benefit most was the constraint on total travel time in which the net benefit was reduced by 90%, but this does depend on the threshold chosen for each constraint.

Tests were also conducted on different variants of the penalty based algorithm. These include the static, dynamic, and self-adaptive penalty based method. All three methods tested in this chapter successfully found the same final solution. As expected, different penalty based methods behave differently due to their underlying formulation. Using the static penalty based method resulted in very flat trends of the average net benefit and net revenue. This is reasonable given the underlying constant function of the static penalty. Intuitively, the behaviours of the dynamic and self-adaptive penalty methods were different from the static one. The intention of these methods is to allocate more effort in optimising the main objective function in early generations (using a low penalty term), and then gradually increasing the penalty as the number of generations increases. Again, the GA based algorithm was found to be flexible enough to take account of both the topological constraint (closed cordon formation) and the additional outcome constraints of the design.

The last requirement left for the development of the optimisation algorithm for the practical cordon design is the treatment of the multiobjective design. Based on the review, four different types of preference expression were found including weight, priority, goal/target, and linguistic ranking. The weighting approach (or utility based approach) is probably the most widely adopted method due to its simplicity. The linguistic ranking approach is the most recent approach which has become feasible following developments in the area of fuzzy reasoning and evolutionary optimisation.

Regardless of the form of the preference, different stages of preference articulation can be adopted. Three possible approaches for deciding on the preference were discussed in the chapter including the priori, posterior, and progressive preference articulation approaches. The combination between the priori preference articulation and the weighting or utility based approach is probably the simplest method for solving the multiobjective problem. This is also the case in transport. However, several drawbacks of this approach were discussed in the chapter. The main problem seems to be the difficulty in making the judgement on the preference prior to obtaining the actual trade-off information between different objectives. Recent developments in evolutionary optimisation or in particular GA allow a better treatment to the multiobjective optimisation problem.

Thus, the GA based algorithm was developed to solve the multiobjective road pricing design with the posterior and progressive articulation approach. The classical concept of Pareto front was adopted to define the fitness of the chromosomes in GA. A solution was considered to be a Pareto solution if none of the other solutions achieves better level of all objectives considered. The algorithm developed follows the framework of the Non-dominated Sorting Genetic Algorithm-II (NSGA-II). The NSGA-II was integrated with the GA-AS method developed earlier. In addition, the NSGA-II was modified to allow for user interruption under the framework of progressive preference articulation.

Four tests with two objectives and one test with three objectives were the objects of the methodological experiment. The NSGA-II with GA-AS was successfully applied to all of these tests to generate a set of non-dominated solutions. Despite the success of the tests, the results revealed a possible weakness of the NSGA-II in which the set of non-dominated solutions identified by the algorithm may not be well spread over the Pareto front. This is due to the complexity of the multiobjective problem and limited computation resource. The progressive preference articulation based approach was adopted to improve the algorithm. In a real problem, the decision maker may not be interested in all parts of the Pareto front. Thus, by allowing the decision maker to express his interest in some particular area of the Pareto front progressively, the algorithm can then allocate more effort on finding a non-dominated solution in that particular area.

10.5 SOME POLICY IMPLICATIONS

The spin-off results from the numerical tests with the algorithms developed in the study also suggest a number of interesting policy implications for road pricing scheme design.

Firstly, the Edinburgh network demonstrated the importance of the toll level and toll location on the performance of the road pricing scheme. The implementation of the optimised toll level around the inner1 and outer cordons of the Edinburgh network can increase the benefit by around 70% compared to the uniform toll. Similarly, the OPC1 cordon found by GA-AS offers about 80% higher benefit compared to the inner2 cordon. In addition, allowing a second cordon, D-OPC, brought a substantial increase in the benefit of the scheme (112% higher than the benefit from the OPC1). With the current practice of the determination of the toll level and location based on the acceptability and professional judgment, a significant amount of benefit of road pricing schemes may be lost.

Secondly, the additional constraints included into the analysis can potentially reduce the benefits of a road pricing scheme. From the tests, the constraints on the guaranteed level of revenue generated, equity impact, total travel time, and total travel distance all resulted in the reduction of the benefit of the cordon scheme in the Edinburgh network.

Obviously, from the policy point of view this presents a political dilemma. On one hand, the inclusion of these constraints will raise the public acceptability of the scheme. On the other hand, the reduction of the benefit of the scheme may question the justification of the implementation of road pricing.

Thirdly, different objectives of road pricing require different characteristics of the charging cordon design. For the objective of social welfare improvement, the cordon should try to concentrate on tackling the trips generating the highest external cost or congestion in the network. The objective of revenue maximisation requires the cordon to charge as many trips as possible with a very high toll so the cordon should consist of many crossing points and offer no alternative routes to avoid the toll. On the other hand, the equity cordon aims to distribute the benefit of the road pricing scheme equally amongst different groups of users from different O-D pairs. Thus, the equity cordon should cover a wide area and apply a toll level that provides a similar level of benefit to all O-D pairs (usually with a low toll level). From these characteristics, it is certain that there is a conflict between different objectives of the road pricing scheme. The decision maker should ensure the understanding of the public and stakeholder groups of the trade-off between different objectives.

10.6 RECOMMENDATIONS FOR FURTHER RESEARCH

During the course of this research, several issues have been identified which require specific further investigation in order to advance the algorithm for designing the optimal road pricing scheme. These are listed below.

Modelling environment and assumptions

The model adopted in this thesis is assumed to follow a number of simplistic assumptions including single user class, single mode, single time period, and the Wardrop's user equilibrium concept. In fact, the applicability of the GA based algorithm developed does not depend on these modelling assumptions. The future research should explore the possibility of the extension of the modelling assumption in order to increase the realism of the model. At the initial stage, one could extend the concept of Wardrop's user equilibrium to increase the dimensions of the model. The extension of the model to allow multiple user classes (e.g. by income) will allow a more detailed analysis of the equity impact (vertical equity). Meng *et al* (2002) and Small and Yan (2001) already studied the possible extension of the optimal road pricing design with multiple user classes.

The inclusion of multiple time periods, other modes (e.g. transit and bus) and land use response will also increase the possible response of the users to the road pricing scheme. Some research has already been done on this issue but at a more strategic model level (Minken *et al*, 2003). Last but not least, the current model adopted is under the assumption of separable link cost function. This assumption may not be realistic for some urban network where the congestion is mainly caused by junction. Thus, it is also important to attempt to link the GA-AS algorithm with model with junction interaction.

The other direction of the modeling improvement is the deviation from the concept of deterministic user equilibrium. Different modeling paradigms can be adopted to better represent the users' behaviours in the network including the stochastic equilibrium model or learning process. Some researchers have already started exploring the development in this area (See for example Sumalee *et al*, 2004).

Extension of the method to cope with a variety of designs

In relation to the first research issue, with more complex model adopted a greater variety of road pricing schemes can be considered. With the model of time period choice, the cordon location may be optimised by time period (of course this should be considered with the cost of the scheme). Similarly, the dis-aggregation of the user class and vehicle class enables the analysis of the charge by user type and could include the possibility of exemption. In fact, the issue of exemption could also be analysed with the current setting of the model if the exemption is defined geographically.

The increase in the dimension of the scheme design is also envisaged as an important research area. The designer should be able to identify the possible use of screen lines and spurs. Similarly, the other key important design option is the integration between the charging cordon and other traffic management measures, e.g. area access control or traffic network control. The cordon scheme may not just use the toll points to form a cordon. At some part of the network, a road could be closed to disallow the access or controlled by some traffic management measure.

Adding time dimension and evolution process

Following the finding from the surveys with the UK local authorities, the design of a road pricing scheme should be considered as an evolutionary process. The initial scheme should be a very simple and acceptable scheme. Then, the scheme can be gradually adjusted to better achieve its goals. This is exemplified by the adjustment of the scheme in Singapore and the proposed change of the scheme in London. With the extension of the time-horizon dimension into the modelling environment (i.e. includes land use response and some form of time-lag response), the scheme can be optimised with different levels of constraints at different time horizons with the aim to optimise the design over the whole time-horizon rather than just concentrate on a particular time-horizon. This extension will definitely need some analysis of the investment decision (cash-flow analysis) to trade-off the benefit at the present time and in the future and can be linked to other issues such as intra-generational equity and sustainability.

Solving MPEC is still a challenging and important task

Although a number of studies have been done on the area of MPEC, the development of a practical and efficient solution algorithm is still challenging. In light of the recent development of a variety of new ideas (e.g. cutting plane algorithm, marginal function based algorithm, non-smooth optimisation), it is envisaged that the development of a practical algorithm for solving a large scale optimal toll problem can be achieved in the near future.

However, there still exists the most complex problem which is the mixed 1-0 optimisation problem of the MPEC. This problem is indeed the optimal toll location design with optimal variable tolls. Some experiments (Shepherd and Sumalee, 2004) which were not reported in this study show a promising result of the application of GA to this problem. However, the main problem with GA is the computational time, and it will be useful if there are further developments in integrating a better heuristic algorithm with GA for solving the optimal toll location problem.

Development of interactive tool for trading-off process

On the issue of multiobjective design, the development of a GA based method in this thesis demonstrates a new way to the appraisal process for a transport project. The development of the algorithm with the linguistic preference may enhance the decision making process of the road pricing design given that some of the criteria is very subjective (e.g. equity). In addition, the advance of the progressive preference articulation method could be potentially useful for the discussion of the scheme design in the early stage of the development.

Integrated strategy and revenue recycling

Road pricing is not the only transport policy. In fact, road pricing is considered as a central part of the idea of an integrated transport policy. In reality, a number of transport policy instruments will be implemented alongside road pricing in order to support or alleviate the impact of road pricing (e.g. improvement of public transport, new road construction, traffic management, etc.). Thus, the optimisation of road pricing as a sole policy may not be sufficient to ensure the optimality of the overall transport system. With the extension of the modelling environment mentioned earlier, one could attempt to include other types of policy into the optimisation process simultaneously with the design of the road pricing scheme.

A related issue to the integration of the transport policy is the recycling of revenue. A crucial assumption made that justifies the benefit of road pricing is the way in which the revenue is recycled back to the user in a perfect manner. This is a crucial assumption both for the analysis of the benefit of road pricing and the issue over the equity impact as discussed earlier. If the approach to spend the revenue collected from the scheme is optimised simultaneously with the design of the road pricing, it may be possible to calculate the true benefit of the road pricing with different assumptions about the use of revenues.

Of course, this is a very complex economic issue since one could argue that there are many different ways the revenue can be recycled even outside the transport area (e.g. through health or educational service). However, recent research suggests that the public would accept road pricing more if the revenue is to be spent to improve the transport system.

Dealing with variability of the modelling parameters and real world uncertainty

One of the major flaws of the analysis presented in this study is the choice of the modelling parameters (e.g. value of time, demand, or demand elasticity). The outcome design of the scheme may vary significantly with the modelling parameters adopted. This study did not explore this angle of the research but it recognises this important issue. One possible framework that can be adopted to reduce the doubt over the result of the optimisation process is the stochastic optimisation framework. The parameters involved in the model can be defined with some possible error (i.e. define the parameters as a statistical distribution) instead of simply adopting one value. Then, the optimisation process can be used to define the optimal cordon scheme under the stochastic parameters using the idea of stochastic optimisation.

The other different angle also involved with the uncertainty of the result is the inherit uncertainty in the future prediction and circumstance. Apart from the uncertainty of the modelling parameters, there exists in the real world some uncertainty, especially with the future forecast. This may involve uncertainty of the demand forecast, exchange rate, economy, etc. These uncertainties should be included into the analysis of optimal road pricing design. Of course, in order to analyse the time-related uncertainty issue the model needs to be equipped with the time-dimension. This issue was discussed earlier.

Finally, with the new framework of the stochastic programming a new type of scheme evaluation may need to be investigated. The traditional framework from welfare economics may not conform to the new stochastic modelling framework. The topic like Value at Risk should be imposed upon the evaluation of the scheme benefit. The decision maker may also need to trade-off between the expected benefit and risk of loss of the benefit.

Multioperators case and tri-level optimisation problem

Last but not least, the analysis conducted in this study mainly involved two types of actors, the leader and followers. The leader in our analysis can be considered as the decision maker of the city who wishes to pursue the idea of road pricing. The followers, as explained earlier, are the users of the network. In reality, the space of the effect of road pricing is not only limited to the space in the model. At the city level, the decision of one city may influence the decision of another nearby city. In this setting, multi leaders may exist in the Stackelberg game. The leaders, who are the city managers, may compete with each other to ensure the development of their own cities. For instance, if the city of Edinburgh decides to implement a road user charging scheme, a competing city (in economic terms) may decide to adopt a strategy to persuade the existing businesses in Edinburgh to relocate to that city.

A similar setting can also be observed where there exists the legal owner of different parts of the network. The city council may be responsible for most of the local roads in the network but the government or private agencies may be responsible for operation of the other trunk roads. In this case, different operators may try to optimise the operation (including the toll) for their parts of the network to maximise their own objectives. Obviously, the decision of one operator will affect the others, so this problem can also be formed as a problem with multi-leaders.

The problem of the multi-leaders with equilibrium constraint has been studied in the field of Equilibrium Programming with Equilibrium Constraint (EPEC). The extension of this problem can be made further by introducing the decision maker at the top of the hierarchy (e.g. central government). With this setup, the government, who is the highest leader in this game, can use this Tri-level problem to analyse the best policy to control the development of road pricing in the whole country making sure the best compromise and co-ordination between different cities. Both EPEC and Tri-level problems are very complex problems and are definitely worthy of further research.

Bibliography

Aashtiani, H. Z., and Magnanti, T. L. (1981). "Equilibria on a congested transportation network." *SIAM Journal on Algebraic and Discrete Methods*, 2, 213-226.

Abdulaal, M., and LeBlanc, L. J. (1979). "Continuous equilibrium network design models." *Transportation Research Part B: Methodological*, 13(1), 19-32.

Allenson, R. (1992). "Genetic algorithms with gender for multi-function optimisation." *Technocal Report EPCC-SS92-01*, Edinburgh Parallel Computing Centre, Edinburgh, Scotland.

Anderson, D., and Mohring, H.(1995) "Congestion costs and congestion pricing." *Proc. of Conference on Congestion Pricing*, Beckman Center, Irvine, California.

Andreani, R., and Martinez, J. M. (2001). "On the solution of mathematical programming problems with equilibrium constraints by means of nonlinear programming algoritms." *Mathematical Methods of Operations Research*, 54(3), 345-358.

Anitescu, M. (2000). "On solving mathematical programs with complementarity constraints as nonlinear programs." *Preprint ANL/MCSP864-1200*, MCS Division, Argonne National Laboratory, Argonne, IL, USA.

Armstrong-Wright, A. T. (1986). "Road pricing and user restraint: Opportunities and constraints in developing countries." *Transportation Research Part A: General*, 20(2), 123-127.

Arnott, R., De Palma, A., and Lindsey, R. (1990a). "Economics of a bottleneck." *Journal of Urban Economics*, 27, 111-130.

Arnott, R., De Palmer, A., and Lindsey, R. (1990b). "Departure Time and Route Choice for the Morning Commute." *Transportation Research, Part B*, 28(24 Refs).

Bard, J. F. (1983). "An efficient point algorithm for a linear two-stage optimization problem." *Operations Research*, 31(4), 670-684.

Bard, J. F. (1988). "Convex Two-Level Optimization." *Mathematical Programming*, 40, 15-27.

Bard, J. F. (1998). *Practical Bilevel Optimization: Algorithms and Applications*, Kluwer Academic Publishers, Boston.

Bartley, B. (1995). "Mobility impacts, reactions and opinions. Traffic demand management options in Europe: The MIRO Project." *Traffic Engineering And Control*, 36, 596-603.

Bazaraa, M. S., Sherali, H. D., and Shetty, C. M. (1993). *Nonlinear programming: theory and algorithms*, John Wiley & Sons, Inc., New York.

Bean, J. C. (1994). "Genetics and random keys for sequencing and optimization." *ORSA Journal on Computing*, 6(2), 154-160.

Beckmann, M. J., McGuire, C. B., and Winsten, C. B. (1956). *Studies in the Economics of Transportation*, Yale University Press, New Haven, Conn.

Ben-Ayed, O., and Blair, I., Charles E. (1990). "Computational difficulties of bilevel linear programming." *Operations Research*, 38, 556-559.

Ben-Ayed, O., Boyce, D. E., and Blair, I., Charles E. (1988). "A general bilevel linear programming formulation of the network design problem." *Transportation Research Part B: Methodological,* 22(4), 311-318.

Benson, H. Y., Shanno, D. F., and Vanderbei, R. J. (2004). "Interior-point methods for nonconvex nonlinear programming." *Mathematical Programming A*, 99, 35-48.

Bergendorff, P., Hearn, D. W., and Ramana, M. V. (1997). "Congestion toll pricing of traffic networks." Lecture Notes in Economics and Mathematical Systems No.450: Network Optimization, P. M. Pardalos, D. W. Hearn, and W. W. Hager, eds., Springer, Berlin, 51-71.

Bialas, W. F., and Karwan, M. H. (1982). "On Two-Level Optimization." *IEEE Transactions on Automatic Control,* 27(1), 211-214.

Blythe, P. T., and Burden, M. J. J.(1996) "The technical and institutional issues associated with the enforcement of a multi-land debiting system." *Proc. of IEE Colloquium on Camera Enforcement of Traffic Regulations.*

Bonsall, P., and Cho, H. J.(1999) "Travellers' response to uncertainty: the particular case of drivers' response to imprecisely known tolls and charges." *Proc. of European Transport Conference: Seminar F,* UK.

Bonsall, P. W., and Palmer, I. (1997). "Do time-based road-user charges induce risk-taking? - Results from a driving simulator." *Traffic Engineering And Control,* 38(4), 200-204.

Boot, J., Boot, P., and Verhoef, E. T.(1999) "The long road towards the implementation of road pricing: the Dutch experience." *Proc. of The ECMT/OECD workshop on managing car use for sustainable urban travel,* Dublin, Ireland.

Borins, S. F. (1988). "Electronic road pricing: An idea whose time may never come." *Transportation Research Part A: General,* 22(1), 37-44.

Bristow, A. L., and Nellthorp, J. (2000). "Transport project appraisal in the European Union." *Transport Policy,* 7(1), 51-60.

Camponogara, E., and Talukdar, S. N. (1997). "A Genetic Algorithm for Constrained and Multiobjective Optimization." 3rd Nordic Workshop on Genetic Algorithms and Their Applications (3NWGA), J. T. Alander, ed., University of Vaasa, Vaasa, Finland, 49-62.

Candler, W., and Norton, R. (1977). "Multilevel programming." *Technical Report 20,* World Bank Development Research Center, Washington D.C.

Carey, M., Ge, Y. E., and McCartney, M. (2003). "A Whole-Link Travel-Time Model with Desirable Properties." *Transportation Science,* 37(1), 83-96.

Catling, I.(2000) "Road user charging using vehicle positioning systems." *Proc. of Road transport information and control,* London, UK.

CfIT. (2002). "Paying for road use." Commission for integrated transport, London, UK.

Chankong, V., and Haimes, Y. Y. (1983). *Multiobjective Decision Making Theory and Methodology,* North-Holland, New York.

Charles, P. (2001). "Begging to differ: Tolling interoperability in Australia." TOLLtrans, 64-67.

Chen, M., Bernstein, D. H., Chien, S. I. J., and Mouskos, K. C. (1999). "Simplified Formulation of the Toll Design Problem." *Transportation Research Record,* 95(3 Fig., 4 Tab., 24 Ref).

Chen, O. J., and Bernstein, D. (1995). "AM/PM congestion pricing with a single toll palza." *Transportation Research Record*, 1498, 23-31.

Chen, S. J., Hwang, C. L., and Hwang, F. P. (1992). *Fuzzy Multiple Attribute Decision Making*, Springer-Verlag, Berlin, Germany.

Chen, Y., and Florian, M. (1995). "The nonlinear bilevel programming problem: Formulation, regularity and optimality conditions." *Optimization*, 32(193-209).

Chin, K. K.(2002) "Road pricing Singapore's experience." *Proc. of IMPRINT-EUROPE Thematic Network: "Implementing Reform on Transport Pricing: Constraints and solutions: learning from best practice"*, Brussels, Belgium.

Chu, X. (1995). "Endogenous trip scheduling: the Henderson approach reformulated and compared with the Vickrey approach." *Journal of Urban Economics*, 37, 324-343.

Clark, J. (2000). "Sky high tolling." ITS International.

Clarke, F. H. (1983). *Optimization and nonsmooth analysis*, John Wiley & Sons, New York, US.

Clegg, J., Smith, M., Xiang, Y., and Yarrow, R. (2001). "Bilevel programming applied to optimising urban transportation." *Transportation Research Part B: Methodological*, 35(1), 41-70.

Coello, C. A. C. (1999a). "A Survey of Constraint Handling Techniques used with Evolutionary Algorithms." *Lania-RI-99-04*, Laboratorio Nacional de Inform'atica Avanzada, Xalapa, Veracruz, M'exico.

Coello, C. A. C.(1999b) "The use of a multiobjective optimization technique to handle constraints." *Proc. of the Second International Symposium on Artificial Intelligence (Adaptive Systems)*, La Habana, Cuba, 251-256.

Coello, C. A. C. (2000a). "Constraint-handling using an evolutionary multiobjective optimization technique." *Civil Engineering and Environmental Systems*, 17, 319-346.

Coello, C. A. C. (2000b). "Treating Constraints as Objectives for Single-Objective Evolutionary Optimization." *Engineering Optimization*, 32(3), 275-308.

Coello, C. A. C. (2002). "Theoretical and numerical constraint-handling techniques used with evolutionary algorithms: a survey of the state of the art." *Computer Methods in Applied Mechanics and Engineering*, 191(11-12), 1245-1287.

Coello, C. C. (2000c). "Use of a Self-Adaptive Penalty Approach for Engineering Optimization Problems." *Computers in Industry*, 41(2), 113-127.

Cohen, Y. (1987). "Commuter welfare under peak period congestion tolls: Who gains and who loses?" *International Journal Of Transport Economics*, 14(3), 239-266.

Coit, D. W., Smith, A. E., and Tate, D. M. (1996). "Adaptive Penalty Methods for Genetic Optimization of Constrained Combinatorial Problems." *INFORMS Journal on Computing*, 8(2), 173-182.

Commission of European Community. (1999). "DIRECTIVE 1999/62/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 17 June 1999 on the charging of heavy goods vehicles for the use of certain infrastructures." Commission of European Community.

Commission of European Community. (2001). "Europe Transport Policy for 2010: Time to decide." Commission of European Community, Brussels, Belgium.

Confederation of British Industry. (1994). *Missing Links: Settling National Transport Priorities*, London.

Conn, A. R., Gould, N. I. M., and Toint, P. L. (1996). "Numerical experiments with the lancelot package (Release A) for large-scale nonlinear optimization." *Mathematical Programming A*, 73(1), 73-110.

Cree, N. D., Maher, M. J., and Paechter, B. (1998). "The continuous equilibrium optimal network design problem: a genetic approach." Transportation Networks: Recent Methodological Advances, M. G. H. Bell, ed., Pergamon, Amsterdam.

Crossley, W. A., and Williams, E. A.(1997) "A Study of Adaptive Penalty Functions for Constrained Genetic Algorithm Based Optimization." *Proc. of AIAA 35th Aerospace Sciences Meeting and Exhibit*, Reno, Nevada.

Cvetkovic, D., and Parmee, I. C. (2002). "Preferences and their application in evolutionary multiobjective optimization." *IEEE Transactions on Evolutionary Computation*, 6(1), 42-57.

Dafermos, S. (1980). "Traffic equilibrium and variational inequalities." *Transportation Science*, 14, 42-54.

Davis, L. (1987). *Genetic Algorithms and Simulated Annealing*, Pitman, London.

Davis, L. (1991). *Handbook of Genetic Algorithms*, Van Nostrand Reinhold, New York.

Davis, L., and Steenstrup, M. (1987). "Genetic Algorithms and Simulates Annealing: An overview." Genetic Algorithms and Simulated Annealing, L. Davis, ed., Morgan Kaufmann Publishers, Los Altos, California.

Dawson, J. A. L., and Brown, F. N. (1985). "Electronic Road Pricing in Hong Kong. 1. A Fair Way to Go?" *Traffic Eng Control*, 26(11), p522-5.

Dawson, J. A. L., and Catling, I. (1986). "Electronic road pricing in Hong Kong." *Transportation Research Part A: General*, 20(2), 129-134.

De Jong, K. A. (1975). "Analysis of the Behavior of a Class of Genetic Adaptive Systems," PhD,Department of Computer and Communication Science, University of Michigan, Ann Arbor, MI.

Deb, K. (2000). "An Efficient Constraint Handling Method for Genetic Algorithms." *Computer Methods in Applied Mechanics and Engineering*, 186(2/4), 311-338.

Deb, K. (2001). *Multi-Objective Optimization using Evolutionary Algorithms*, John Wiley & Sons, Ltd., New York.

Deb, K., Agrawal, S., Pratap, A., and Meyarivan, T.(2000) "A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II." *Proc. of the Parallel Problem Solving from Nature VI (PPSN-VI)*, 849-858.

Dempe, S. (2002). *Foundations of Bilevel Programming*, Kluwer Academic Publishers, Boston.

Dempe, S. (2003). "Annotated Bibliography on Bilevel Programming and Mathematical Programs with Equilibrium Constraints." *Optimization*, 52, 333-359.

Dem'yanov, V. F., and Malozemov, V. N. (1990). *Introduction to minmax*, Dover, New York.

Department for Transport. (2000). *Transport 10 Year Plan 2000*, HMSO, London.

Department for Transport. (2001). "Perceptions of and attitudes to congestion." Department for Transport, London.

Department for Transport and Regional Services. (2002). "AusLink: Towards the National Land Transport Plan." Ministry for Transport and Regional Services,.

Department of Environment/Scottish Office. (1996). *The United Kingdom National Air Quality Strategy (consultation draft)*, HMSO, London.

Department of Transport. (1989). "Highways Economics Note No.2 in COBA 9 Manual." HMSO, London.

Depuit, J. (1844). "On the measurability of the utility of public works." Transport, D. Murphy, ed., Penguin, London.

DETR. (1998). "Breaking the Logjam: the Government's consultation paper on fighting traffic congestion and pollution through road user and workplace parking charges." Department of Environment, Transport and the Regions, UK.

DETR. (2000). "Guidance on Local Transport Plans." HMSO, London.

Dittmar, H., Frick, K., and Tannehill, D. (1994). "Institutional and political challenges in implementing congestion charging: case study of the San Francisco Bay area." *Transport Research Board Special Report*, 242(2).

Drezner, Z., and Wesolowsky, G. O. (2003). "Network design: selection and design of links and facility location." *Transportation Research Part A: Policy and Practice*, 37(3), 241-256.

Dutch Minister of Transport. (1995). "Contours of implementation of congestion charging (Rekening Rijden)." Abstract of a Letter to Parliament from the Minister of Transport.

Edmunds, T. A., and Bard, J. F. (1991). "Algorithms for Nonlinear Bilevel Mathematical Programs." *IEEE Transactions on systems, man, cybernatics*, 21(1), 83-89.

Eliasson, J., and Lundberg, M. (2003). "Road pricing in urban areas." *136E*, Swedish National Road Administration.

Else, P. (1986). "No entry for congestion taxes." *Transportation Research*, 20A(2), 99-107.

Emberger, G., Shepherd, S. P., and Zhang, X.(2003) "Method to identify optimal land use and transport policy packages: a comparison of CBA and indicator based optimisation." *Proc. of European Transport Conference*, Strasbourg, France.

Facchinei, F., Jiang, H., and Qi, L. (1999). "A smoothing method for mathematical programs with equilibrium constraints." *Mathematical Programming*, 85, 107-134.

Falk, J. E., and Liu, J. (1992). "On Bilevel Programming, Part 1: General Nonlinear Cases." *Mathematical Programming*, 70(1), 47-72.

Ferrari, P. (1995). "Road pricing and network equilibrium." *Transportation Research Part B: Methodological*, 29(5), 357-372.

Ferris, M. C., and Pang, J. S. (1997). "Engineering and economic applications of complementarity problems." *SIAM Review*, 39(4), 669-713.

Fiacco, A. V. (1976). "Sensitivity analysis for non-linear programming using penalty method." *Mathematical Programming*, 10, 287-311.

Fiacco, A. V., and McCormick, G. P. (1968). *Nonlinear Programming: Sequential Unconstrained Minimization Technique*, John Wiley and Sons, New York.

Fischer, A. (1995). "A Newton-type method for positive semidefinite linear complementarity problems." *Journal of Optimization Theory and Applications*, 86, 88-105.

Fisk, C. S. (1984). "Game theory and transportation systems modelling." *Transportation Research Part B: Methodological*, 18(4-5), 301-313.

Fletcher, R., and Leyffer, S. (2002). "Numerical experience with solving MPECs as NLPs." *Numerical Analysis Report NA/210*, Department of Mathematics, University of Dundee, Dundee, UK.

Fletcher, R., Leyffer, S., Ralph, D., and Scholtes, S. (2002). "Local convergence of SQP methods for Mathematical Programs with Equilibrium Constraints." *Numerical Analysis Report NA/209*, Department of Mathematics, University of Dundee, Dundee, UK.

Fonseca, C. M., and Fleming, P. J.(1993) "Genetic Algorithms for Multi-objective Optimization: Formulation, Discussion, and Generalization." *Proc. of the Fifth International Conference on Genetic Algorithms*, San Mateo, Califonia, 416-423.

Fonseca, C. M., and Fleming, P. J. (1998). "Multiobjective Optimization and Multiple Constraint Handling with Evolutionary Algorithms - Part I: A Unified Formulation." *IEEE Transactions on systems, man, cybernatics - Part A: Systems and humans*, 28(1), 26-37.

Foo, T. S. (2000). "An advanced demand management instrument in urban transport: Electronic road pricing in Singapore." *Cities*, 17(1), 33-45.

Fourman, M. P.(1985) "Compation of symbolic layout using genetic algorithms." *Proc. of the First International Conference on Genetic Algorithms*, 141-153.

Fridstrom, L., Minken, H., Moilanen, P., Shepherd, S., and Vold, A. (2000). "Economic and equity effects of marginal cost pricing in transport case studies from thee European cities." *71*, VATT, Helsinki, Finland.

Friesz, T. L., Anandalingam, G., Mehta, N. J., Nam, K., Shah, S. J., and Tobin, R. L. (1993a). "The multiobjective equilibrium network design problem revisited: A simulated annealing approach." *European Journal of Operational Research*, 65(1), 44-57.

Friesz, T. L., Bernstein, D., Smith, T. E., Tobin, R. L., and Wie, B. W. (1993b). "A variational inequality formulation of the dynamic network user equilibrium problem." *Operations Research*, 41, 179-191.

Friesz, T. L., Cho, H. J., Mehta, N. J., Tobin, R. L., and Anandalingam, G. (1992). "A Simulated Annealing Approach to the Network Design Problem with Variational Inequality Constraints." *Transportation Science*, 26, 18-26.

Friesz, T. L., and Harker, P. T. (1983). "Multicriteria spatial price equilibrium network design: Theory and computational results." *Transportation Research Part B: Methodological*, 17(5), 411-426.

Friesz, T. L., and Harker, P. T. (1985). "Properties of the iterative optimization equilibrium algorithms." *Civil Engineering Systems*, 2, 142-154.

Friesz, T. L., Tobin, R. L., Cho, H. J., and Mehta, N. J. (1990). "Sensitivity analysis based heuristic algorithms for mathematical programs with variational inequality constraints." *Mathematical Programming*, 48, 265-284.

Fukushima, M., Luo, Z. Q., and Pang, J. S. (1998). "A Globally Convergent Sequential Quadratic Programming Algorithm for Mathematical Programs with Linear Complementarity Constraints." *Computational Optimization and Applications*, 10, 5-34.

Gen, M., and Cheng, R. (2000). *Genetic Algorithms & Engineering Optimization*, John Wiley and Sons, Inc., Chichester, UK.

Gen, M., Cheng, R., and Oren, S. S. (2001). "Network design techniques using adated genetic algorithms." *Advances in engineering software*, 32, 731-744.

Gerrard, W. (2000). "Traffic demand management in three historical cities: results of a multivariate analysis of business attitudes." *WP 552*, Institute for Transport Studies, University of Leeds, Leeds, UK.

Gill, P. E., Murray, W., and Saunders, M. A. (1997). "SNOPT: An SQP algorithm for large-scale constrained optimization." Department of Mathematics, University of Califonia, San Diego, San Diego.

Giuliano, G. (1992). "An assessment of the political acceptability of congestion pricing." *Transportation*, 19, 335-358.

Giuliano, G. (1994). "Equity and fairness considerations of congestion pricing. Curbing Gridlock: Peak-Period Fess to Relieve Traffic Congestions, Special Report 242, Volume 2." Transportation Research Board, National Research Council, USA.

GLA. (2001). "The Mayor's Transport Strategy: Draft for public consultation." Greater London Authority, London.

Glazer, A., and Niskanen, E. (1993). "Parking fees and congestion." *Regional Science and Urban Economics*, 22, 123-132.

Glover, F. (1989). "Tabu Search-part I." *ORSA Journal on Computing*, 1, 190-206.

Goh, M. (2002). "Congestion management and electronic road pricing in Singapore." *Journal of Transport Geography*, 10(1), 29-38.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, Reading, Massachusetts.

Goldberg, D. E. (2002). *The Design of Innovation: Lesson from and for Competent Genetic Algorithms*, Kluwer Academic Publishers, Dordretch, The Netherlands.

Gomez, A., Fernandez, Q. I., De la Fuente Garcia, D., and Garcia, P. J. (2003). "Using genetic algorithms to resolve layout problems in facilities where there are aisles." *International Journal of Production Economics*, 84(3), 271-282.

Gomez-Ibanez, J. A. (1992). "The political economy of highway tolls and congestion pricing." Exploring the Role of Pricing as a Congestion Management Tool, Federal Highway Administration, ed., FHA, Washington, USA.

Gomez-Ibanez, J. A., and Small, K. A. (1994). *Road Pricing for Congestion Management: A Survey for International Practice*, National Academy Press, Washington D.C.

Goodwin, P. B. (1989). "The rule of three: a possible solution to the political problem of competing objectives for road pricing." *Traffic Engineering And Control*, 30(10).

Greater London City Council. (1974). "A Study of Supplementary Licensing." London GLC, London.

Guillermo Jordan, J., Soriano, F., Graullera, D., and Martin, G.(2001) "A comparison of different technologies for EFC and other ITS applications." *Proc. of IEEE Intelligent Transportation Systems Conference*, Oakland, USA.

Hadji-Alouane, A. B., and Bean, J. C. (1997). "A genetic algorithm for multiple-choice integer program." *Operations Research*, 45, 92-101.

Haimes, Y. Y., Lasdon, L. S., and Wismer, D. A. (1971). "On a bicriteria formulation of the problems of integrated system identification and system optimization." *IEEE Transactions on Systems, man, cybernatics*, 1(3), 296-297.

Halden, D. (2003). "Using accessibility measures to integrate land use and transport policy in Edinburgh and the Lothians." *Transport Policy*, 9, 313-324.

Hansen, B., Jaumard, B., and Savard, G. (1992). "New branch-and-bound rules for linear bilevel programming." *SIAM Journal on Scientific and Statistical Computing*, 13, 1194-1217.

Hansen, J. V. (2004). "Genetic search methods in air traffic control." *Computers & Operations Research*, 31(3), 445-459.

Harker, P. T., and Pang, J. S. (1988). "On the existence of optimal solutions to mathematical programs with equilibrium constraints." *Operations Research Letters*, 7, 61-64.

Harrison, B. (1986). "Electronic road pricing in Hong Kong: estimating and evaluating the effects." *Traffic Engineering And Control*, 27(1), 13-18.

Harrison, W. J., Pell, C., Jones, P. M., and Ashton, H. (1986). "Some advances in model design developed for the practical assessment of road pricing in Hong Kong." *Transportation Research Part A*, 20(2), 135-143.

Harsman, B., Padam, S., and Wijkmark, B. (2000). "Ways and means to increase the acceptance of urban road pricing." *Deliverable 4, EU-project PRIMA*, the European Commission, 4th Framework Transport RTD.

Hearn, D. W., Lawphongpanich, S., and Nguyen, S. (1984). "Convex programming formulations of the asymmetric traffic assignment problem." *Transportation Research Part B: Methodological*, 18(4-5), 357-365.

Hearn, D. W., and Ramana, M. V. (1998). "Solving congestion toll pricing models." Equilibrium and Advanced Transportation Modeling, P. Marcotte and S. Nguyen, eds., Kluwer Academic Publishers, Boston, 109-124.

Hearn, D. W., and Yildirim, M. B. (2002). "A Toll Pricing Framework for Traffic Assignment Problems with Elastic Demand." Transportation and Network Analysis: Current Trends, M. Gendreau and P. Marcotte, eds., Kluwer Academic Publishers, Dordrecht.

Hearn, D. W., Yildirim, M. B., Ramana, M. V., and Bai, L.(2001) "Computational Methods for Congestion Toll Pricing Models." *Proc. of IEEE Intelligent Transportation System*, Okland.

Henderson, J. V. (1974). "Road congestion pricing: a reconsideration of pricing theory." *Journal of Urban Economics*, 1, 346-365.

Ho, H. W., Wong, S. C., Yang, H., and Loo, B. P. Y.(2004) "Cordon-base congestion pricing in a continuum traffic equilibrium system." *Proc. of The 10th World Conference on Transport Research Society*, Istanbul, Turkey.

Hoffmeister, F., and Sprave, J.(1996) "Problem-independent handling of constraints by use of metric penalty functions." *Proc. of the Fifth Annual Conference on Evolutionary Programming (EP'96)*, 289-294.

Holland, E. P., and Watson, P. L. (1978). "Traffic restraint in Singapore: measuring the impacts of area license scheme." *Traffic Engineering And Control*, 19, 14-17.

Homaifar, A., Lai, S. H. Y., and Qi, X. (1994). "Constrained optimization via genetic algorithms." *Simulation*, 62(4), 242-254.

Horn, J., and Nafpliotis, N. (1993). "Multiobjective Optimization using the Niched Pareto Genetic Algorithms." *Technical Report IlliGA1 Report 93005*, University of Illinois Urbana-Champaign, Urbana, Illinois, USA.

House of Commons. (1990). "Roads for the future. First Report of the Transport Committee. Session 1989/90." *HC 198-I.*, HMSO, London.

Hwang, C. L., and Masud, A. S. M. (1979). *Multiple Objective Decision Making - Methods and Applications*, Spinger-Verlag, Berlin, Germany.

Hwang, K. Y., Son, B., and Eom, J. K. (1999). "Effect of congestion pricing at the Namsan tunnels in Seoul." *Journal of the Eastern Asia Society for Transportation Studies*, 3(4).

Hyman, G., and Mayhew, L. (2002). "Optimizing the benefits of urban road user charging." *Transport Policy*, 9, 189-207.

Ignizio, J. P. (1976). *Goal Programming and Extensions*, Lexington Books, Lexington, MA.

Ison, S. (2000). "Local authority and academic attitudes to urban road pricing: a UK perspective." *Transport Policy*, 7(4), 269-277.

Jacques, P., Mourad, S., and Bertrand, M. (1997). "GA Multiple Objective Optimization Strategies for Electromagnetic Backscattering." Genetic Algorithms and Evolution Strategies in Engineering and Computer Science. Recent Advances and Industrial Applications, D. Quagliarella, P. Jacques, C. Poloni, and G. Winter, eds., John Wiley and Sons, West Sussex, England, 225-243.

Jaensirisak, S., May, A. D., and Wardman, M. (2003). "Acceptability of road user charging: the influence of selfish and social perspectives." Acceptability of Transport Pricing Strategies, B. Schlag, ed., Elsevier, Oxford.

Jaensirisak, S., Wardman, M., and May, A. D. (2004). "Explaining variations in public acceptability of road pricing schemes." *Transportation Research A*, forthcoming.

Jang, Y.-J., Jang, S.-Y., Chang, B.-M., and Park, J. (2002). "A combined model of network design and production/distribution planning for a supply network." *Computers & Industrial Engineering*, 43(1-2), 263-281.

Jeroslow, R. G. (1985). "The polynomial hierarchy and a simple model for competitive analysis." *Mathematical Programming*, 32, 146-164.

Jiang, H., and Ralph, D. (1997). "QPECgen, a MATLAB generator for mathematical programs with quadratic objectives and affine variational inequality constraints." University of Melbourne, Department of Mathematics, Melbourne.

Joines, J., and Houck, C.(1994) "On the use of non-stationary penalty functions to solve nonlinear constrained optimization problmes with GAs." *Proc. of the First IEEE Conference on Evolutionary Computations*, 579-584.

Jones, P. (1998). "Urban road pricing: public acceptability and barriers to implementation." Road Pricing, Traffic Congestion and the Environment, K. J. Button and E. T. Verhoef, eds., Edward Elgar Publishing Limited, Cheltenham, UK.

Jones, P.(2002) "Addressing equity concerns in relation to road user charging." *Proc. of Conference on Acceptability of Transport Pricing Strategies*, Dresden, Germany.

Jones, P., and Hervik, A. (1992). "Restraining car traffic in European cities: an emerging role for road pricing." *Transportation Research Part A*, 26(2), 133-145.

Kanzow, C. (1996). "Some noninterior continuation methods for linear complementarity problems." *SIAM Journal of Matrix Analysis and Application*, 17, 851-868.

Keeney, R. L., and Raiffa, H. (1976). *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*, Wiley, New York.

Kirkpatrick, S., Gelatt, C. D. J., and Vecchi, M. P. (1983). "Optimization by simulated annealing." *Science*, 220, 671-680.

Knight, F. H. (1924). "Some fallacies in the interpretation of social cost." *Quaterly Journal of Economics*, 38, 582-606.

Knowles, J. D., and Corne, D. W. (2000). "Approximating the non-dominated front using Pareto archived evolutionary strategy." *Evolutionary Computation*, 8(2), 149-172.

Kolstad, C. D., and Lasdon, L. S. (1990). "Derivative Evaluation and Computational Experience with Large Bilevel Mathematical Programs." *Journal of Optimization Theory and Applications*, 65, 485-499.

Koziel, S., and Michalewicz, Z.(1998) "A Decoder-based Evolutionary Algorithm for Constrained Parameter Optimization Problems." *Proc. of the 5th Parallel Problem Solving from Nature (PPSN V)*, Heidelberg, Germany, 231-240.

Kumagai, Y. (2003). "Tolling Technology: Growing in popularity - reviews developments in Japan." ITS International, 47-48.

Kuri Morales, A., and Quezada, C. V.(1998) "A universal eclectic genetic algorithms for constrained optimization." *Proc. of the 6th European Congress on Intelligent Techniques and Soft Computing, EUFIT'98*, 518-522.

Langmyhr, T. (1997). "Managing equity : The case of road pricing." *Transport Policy*, 4(1), 25-39.

Larsen, O., and Ramjerdi, F. (1991). "Road pricing as a means of financing investment in transport infrastructure- the case of Oslo." Institute of Transport Economics, Oslo.

Laumanns, M., Thiele, L., Deb, K., and Zitzler, E. (2002). "Combining Convergence and Diversity in Evolutionary Multiobjective Optimization." *Evolutionary Computation*, 10(3), 263-282.

Lawphongpanich, S., and Hearn, D. W. (1984). "Simplical decomposition of the asymmetric traffic assignment problem." *Transportation Research Part B: Methodological*, 18(2), 123-133.

Lawphongpanich, S., and Hearn, D. W. (2004). "An MPEC Approach to Second-Best Toll Pricing." *Mathematical Programming B*, forthcoming.

Layard, R. (1997). "Cost-Benefit Analysis." Penguin Books, England.

LeBlanc, L. J., and Boyce, D. E. (1986). "A bilevel programming algorithm for exact solution of the network design problem with user-optimal flows." *Transportation Research Part B: Methodological*, 20(3), 259-265.

LeBlanc, L. J., and Boyce, D. E. (1988). "The network design problem: A comment on the paper by Leblanc and Boyce." *Transportation Research Part B: Methodological*, 22(3), 236-237.

Lee Jr., D. B. (2000). "Methods for evaluation of transportation projects in the USA." *Transport Policy*, 7, 41-50.

Legislative Council. (2001). "Electronic road pricing." Legislative council, Hong Kong.

Leung, M., and Liu, L. (1985). "Government caught at the cross-roads." South China Morning Post, Hong Kong, 5.

Leventhal, T., Nemhauser, G. L., and Trotter, L. E. (1973). "A column generation algorithm for optimal traffic assignment." *Transportation Science*, 7, 168-176.

Levy-Lambert, H. (1968). "Tarification des services a qualite variable: application aux peages de circulation." *Econometrica*, 36, 564-574.

Lewis, N. C. (1993). *Road pricing: Theory and practice*, Thomas Telford, London.

Leyffer, S. (2002). "The Penalty Interior Point Method fails to converge for Mathematical Programs with Equilibrium Constraints." *Numerical Analysis Report NA/208*, Department of Mathematics, University of Dundee, Dundee, UK.

Liepins, G. E., and Vose, M. D. (1990). "Representational Issues in Genetic Optimization." *Journal of Experimental and Theoretical Computer Science*, 2(2), 4-30.

Lighthill, M. H., and Whitham, G. B. (1955). "On kinematic waves, II: A theory of traffic flow on long crowded roads." *Proc. of the Royal Society (London)*, A229, 317-345.

Lim, A. C. (2002). "Transportation Network Design Problems: An MPEC Approach,"Departmen of Mathematics, The Johns Hopkins University, Baltimore, Maryland.

Lin, G. H., and Fukushima, M. (2003). "New Relaxation Method for Mathematical Programs with Complementarity Constraints." *Journal of Optimization Theory and Applications*, 118(1), 81-116.

Lis, J., and Eiben, A. E.(1996) "A Multi-Sexual Genetic Algorithm for Multi-objective Optimization." *Proc. of the 1996 International Conference on Evolutionary Computation*, Nagoya, Japan, 59-64.

Liu, L. N., and McDonald, J. F. (1999). "Economic efficiency of second-best congestion pricing schemes in urban highway systems." *Transportation Research Part B*, 33(3), 157-188.

Liu, X., and Sun, J. (2004). "Generalized Stationary Points and an Interior-Point Method for Mathematical Programs with Equilibrium Constraints." *Mathematical Programming*, forthcoming.

Luo, Z. Q., Pang, J. S., and Ralph, D. (1996). *Mathematical programs with equilibrium constraints*, Cambridge University Press, New York.

Marchand, M. (1968). "A note on optimal tolls in an imperfect environment." *Econometrica*, 36, 575-581.

Marcotte, P. (1988). "A note on a bilevel programming algorithm by Leblanc and Boyce." *Transportation Research Part B: Methodological*, 22(3), 233-236.

Marcotte, P., and Marquis, G. (1992). "Efficient implementation of heuristics for the continuous network design problem." *Annals of Operations Research*, 34, 163-176.

Marcotte, P., Savard, G., and Zhu, D. L. (2001). "A trust region algorithm for nonlinear bilevel programming." *Operations Research Letters*, 29(4), 171-179.

Marcotte, P., and Zhu, D. L. (1997). "Exact and inexact penalty methods for the generalized bilevel programming problem." *Mathematical Programming*, 74, 141-157.

May, A. D. (1975). "Supplement licensing: an evaluation." *Traffic Engineering And Control*, 16(4), 162-167.

May, A. D. (1992). "Road pricing: an international perspective." *Transportation*, 19(4), 313-333.

May, A. D., Coombe, D., and Travers, T. (1996). "The London Congestion Charging Research Programme. 5. Assessment of the Impacts." *Traffic Engineering And Control*, 37(6), p403-8.

May, A. D., Liu, R., Shepherd, S. P., and Sumalee, A. (2002). "The impact of cordon design on the performance of road pricing schemes." *Transport Policy*, 9, 209-220.

May, A. D., and Milne, D. S. (2000). "Effects of alternative road pricing systems on network performance." *Transportation Research Part A: Policy and Practice*, 34(6), 407-436.

May, A. D., and Roberts, M. (1995). "The design of integrated transport strategies." *Transport Policy*, 2(2).

May, A. D., Roberts, M., and Mason, P. (1992). "The development of transport strategies for Edinburgh." *Proc. of the Institution of Civil Engineers and Transport*, 95, 51-59.

May, A. D., and Sumalee, A.(2003) "One step forwards, two step back? An overview of road pricing application and research outside the US." *Proc. of International Symposium on Road Pricing*, Key Biscayne, Florida.

Mayeres, I., and Proost, S. (2001). "Tax reform for congestion type of externalities." *Journal of Public Economics*, 79, 343-363.

McDonald, J. F. (1995). "Urban Highway Congestion. An Analysis of Second-Best Tolls." *Transportation*, 22(4), 353-369.

Meng, Q., and Yang, H. (2002). "Benefit distribution and equity in road network design." *Transportation Research Part B: Methodological*, 36(1), 19-35.

Meng, Q., Yang, H., and Bell, M. G. H. (2001). "An equivalent continuously differentiable model and a locally convergent algorithm for the continuous network design problem." *Transportation Research Part B: Methodological*, 35(1), 83-105.

Menon, A. P. G. (2000). "ERP in Singapore - a perspective a year on." *Traffic Engineering And Control*, 41(2).

Michalewicz, Z. (1992). *Genetic Algorithms + Data Structures = Evolution Programs*, Springer-Verlag, New York.

Michalewicz, Z., and Attia, N. F.(1994) "Evolutionary Optimization of Constrained Problems." *Proc. of the 3rd Annual Conference on Evolutionary Programming*, 98-108.

Miettinen, K. (1999). *Nonlinear Multiobjective Optimization*, Kluwer, Boston.

Milne, D. S. (1997). "Modelling the network effects of urban road user charging,"Institute for Transport Studies, University of Leeds, Leeds.

Milne, D. S., Niskanen, E., and Verhoef, E. (2001). "Legal and institutional framework for marginal cost pricing in urban transport in europe." *76*, VATT, Helsinki, Finland.

Ministry of Transport. (1964). "Road pricing: the economic and technical possibilities." HMSO, London.

Ministry of Transport. (1967). "Better Use of Town Roads." HMSO, London.

Ministry of Transport. (1998). "A new deal for Transport White Paper: Better for everyone." HMSO, London.

Minken, H., Jonsson, D., Shepherd, S. P., Jarvi, T., May, A. D., Page, M., Pearman, A., Pfaffenbichler, P., Timms, P., and Vold, A. (2003). "Developing Sustainable Urban Land Use and Transport Strategies: A Methodological Guidebook." *TOI report 619/2003*, Institute of Transport Economics, Oslo.

Morisugi, H. (2000). "Evaluation methodologies of transportation projects in Japan." *Transport Policy*, 7(1), 35-40.

Mun, S., Konishi, K., and Yoshikawa, K.(2001) "Optimal Cordon Pricing." *Proc. of The Ninth World Conference on Transport Research*, Seoul, Korea.

Mun, S. I. (1994). "Traffic jams and the congestion toll." *Transportation Research Part B*, 28(5), 365-375.

Mun, S. I. (1999). "Peak-Load Pricing of a Bottleneck with Traffic Jam." *Journal of Urban Economics*, 46, 323-349.

Muu, L. D., and Quy, N. V. (2003). "A Global Optimization Method for Solving Convex Qudratic Bilevel Programming Problems." *Journal of Global Optimization*, 26, 199-219.

MVA. (1995). *The London Congestion Charging Research Programme: Principle Findings*, HMSO, London.

Nagurney, A. (1993). *Network economics: a variational inequality approach*, Kluwer Academic Publishers, Boston.

Nash, J. F. (1951). "Nocooperative games." *Annals of Mathematics*, 45, 286-295.

Neuenschwander, R. (2000). "Optimal Design of Road Pricing Scheme." *Prima project, the European Union 4th Framework programme*, European Commission.

Nocedal, J., and Wright, S. J. (1999). *Numerical optimization*, Springer-Verlag, Berlin.

Ochieng, W.(2003) "The future for satellite-based charging systems." *Proc. of International Symposium on Congestion Charging*, Imperial College, London.

Odeck, J., and Brathen, S. (2002). "Toll financing in Norway: The success, the failures and perspectives for the future." *Transport Policy*, 9(3), 253-260.

OECD. (1995). *Urban travel and sustainable development*, OECD and EMCT, Paris.

Office for National Statistics. (2001). "Omnibus survey 2001." HMSO, London.

Office for National Statistics. (2003). "Transport trends." HMSO, London.

Oldridge, B.(1990) "Electronic road pricing: an answer to traffic congestion?" *Proc. of Proc of Information Technology and Traffic Management conference*, London.

Ortega, J. M., and Rheinboldt, W. C. (1970). *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York.

Oscar Faber, C. (2001). "Road User Charging Study-West Midlands, Final Report prepared for Birmingham City Council."

Outrata, J. V., Kocvara, M., and Zowe, J. (1998). *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints*, Kluwer Academic Publishers, Boston.

Outrata, J. V., and Zowe, J. (1995). "A numerical approach to optimization problems with variational inequality constraints." *Mathematical Programming*, 68, 105-130.

Pang, J. S., Han, S. P., and Rangaraj, N. (1991). "Minimization of locally Lipschitzian functions." *SIAM Journal of Optimization*, 1, 57-82.

Paredis, J.(1994) "Co-evolutionary constraint satisfaction." *Proc. of the 3rd Confefence on Parallel Problem Solving from Nature*, 46-55.

Patriksson, M. (1994). *The Traffic Assignment Problem: Models and Methods*, VSP, Ultrecht, The Netherlands.

Patriksson, M., and Rockafellar, R. T. (2002). "A Mathematical Model and Descent Algorithm for Bilevel Traffic Management." *Transportation Science*, 36(3), p271-291.

PATS Consortium. (2001). "Recommendations on transport pricing strategies: Final report of the PATS project." European Commission, Brussels.

Phang, S.-Y. (1993). "Singapore's motor vehicle policy: Review of recent changes and a suggested alternative." *Transportation Research Part A: Policy and Practice*, 27(4), 329-336.

Pigou, A. C. (1920). *Wealth and Welfare*, Macmillan, London.

Ramjerdi, F., Minken, H., and Østmoe, K. (2004). "Norwegian urban tolls." Road Pricing: Theory and Evidence, G. Santos, ed., Elsevier, 237-250.

Ran, B., and Boyce, D. E. (1996). *Modelling Dynamic Transportation Networks*, Springer-Verlag, Heidelberg, Germany.

Rasheed, K.(1998) "An Adaptive Penalty Approach for Constrained Genetic-Algorithm Optimization." *Proc. of the Third Annual Genetic Programming Conference*, San Francisco, California, 584-590.

Ray, T.(2002) "Constraint Robust Optimal Design using a Multiobjective Evolutionary Algorithm." *Proc. of the Congress on Evolutionary Computation 2002 (CEC'2002)*, Piscataway, New Jersey, 419-424.

Richardson, J. T., Palmer, M. R., Liepins, G., and Hilliard, M. (1989). "Some Guidelines for Genetic Algorithms with Penalty Functions." Proceeding of the 3rd International Conference on Genetic Algorithms, J. Schaffer, ed., Morgan Kaufmann, CA, USA.

Rietveld, P., and Verhoef, E. T. (1998). "Social feasibility of policies to reduce externalities in transport." Road pricing, Traffic congestion ,and the Environment, K. Button and E. T. Verhoef, eds., Edward Elgar Publishing, Massachusetts, 285-307.

Rom, M. (1994). "The Politics of Congestion Pricing." *Transport Research Board Special Report*, 242(2).

Rosenthal, R. E. (1985). "Principles of multiobjective optimization." *Decision Sciences*, 16(2), 133-152.

SACTRA. (1994). "Trunk Roads and the Generation of Traffic." HMSO, London.

Sayers, T. M., Jessop, A. T., and Hills, P. J. (2003). "Multi-criteria evaluation of transport options--flexible, transparent and user-friendly?" *Transport Policy*, 10(2), 95-105.

Schade, J., and Schlag, B. (2000). "Acceptability of urban transport pricing." *72*, VATT, Helsinki, Finland.

Schade, J., and Schlag, B. (2003). "Acceptability of urban transport pricing strategies." *Transportation Research Part F: Traffic Psychology and Behaviour*, 6(1), 45-61.

Schaffer, J. D.(1985) "Multiple objective optimization with vector evaluated genetic algorithms." *Proc. of the First International Conference on Genetic Algorithms*, 93-100.

Scheel, H., and Scholtes, S. (2000). "Mathematical program with complementarity constraints: Stationarity, optimality, and sensitivity." *Mathematics of Operations Research*, 25, 1-22.

Schlag, B., and Schade, J. (2000). "Public acceptability of traffic demand management and pricing measures in Europe." *Traffic Engineering And Control*, 41(8), 314-318.

Schoenauer, M., and Michalewicz, Z.(1997) "Boundary Operators for Constrained Parameter Optimization Problems." *Proc. of the Seventh International Conference on Genetic Algorithms (ICGA-97)*, San Francisco, California, 322-329.

Schoenauer, M., and Xanthakis, S.(1993) "Constrained GA optimization." *Proc. of the Fifth International Conference on Genetic Algorithms*, 573-580.

Scholtes, S., and Sthör, M. (1999). "Exact Penalization of Mathematical Programs with Equilibrium Constraints." *SIAM Journal on Control and Optimization*, 37(2), 617-652.

Seik, F. T. (1998). "A unique demand management instrument in urban transport: the vehicle quota system in Singapore." *Cities*, 15(1), 27-39.

Sharp, C. H. (1966). "Congestion and welfare: an examination of the case of congestion tax." *Economic Journal*, 76, 806-817.

Sheffi, Y. (1985). *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*, Prentice-Hall, Inc, Englewood Cliffs, New Jersey.

Sheldon, R., Scott, M., and Jones, P.(1993) "London congestion charging: exploratory social research among London residents." *Proc. of 21st PTRS Summer Annual Meeting: Seminar L*, UK, 129-145.

Shepherd, S. P., May, A. D., and Milne, D. S.(2001) "The design of optimal road pricing cordons." *Proc. of the ninth world conference on transport research*, Seoul, Korea.

Shepherd, S. P., and Sumalee, A. (2004). "A Genetic Algorithm Based Approach to Optimal Toll Level and Location Problems." *Networks and Spatial Economics*, 4, 161-179.

Shimizu, K., Ishizuka, Y., and Bard, J. F. (1997). *Nondifferentiable and two-level mathematical programming*, Kluwer Academic, Boston.

Shrank, D., Turner, S., and Lomax, T. (1993). "Estimates of Urban Roadway Congestion, 1990." *Report 1131-5*, Texas Transportation Institute, College Station, Texas.

Small, K. A. (1983). "The incidence of congestion tolls on urban highways." *Journal of Urban Economics*, 13, 90-111.

Small, K. A. (1992). "Using the revenues from congestion pricing." *Transportation*, 19, 359-381.

Small, K. A., and Gomez-Ibanez, J. A. (1998). "Road pricing for congestion management: the transition from theory to policy." Road pricing, traffic congestion, and the environment: issues of efficiency and social feasibility, K. Button and E. T. Verhoef, eds., Edward Elgar Publishing, Massachusetts, 213-246.

Small, K. A., and Yan, J. (2001). "The Value of "Value Pricing" of Roads: Second-Best Pricing and Product Differentiation." *Journal of Urban Economics*, 49(2), 310-336.

Smith, A. E., and Tate, D.(1993) "Genetic Optimization Using a Penalty Function." *Proc. of the Fifth International Conference on Genetic Algorithms (ICGA-93)*, San Mateo, California, 499-503.

Smith, M. J. (1979). "The existence, uniqueness and stability of traffic equilibria." *Transportation Research Part B: Methodological*, 13(4), 295-304.

Srinivas, N., and Deb, K. (1994). "Multiobjective Optimization Using Non-dominated Sorting in Genetic Algorithms." *Evolutionary Computation*, 2(3), 221-248.

Stackelberg, H. V. (1952). *The Theory of Market Economy*, Oxford University Press, Oxford.

Steenbrink, P. A. (1974). *Optimization of Transport Networks*, John Wiley & Sons, London, UK.

Stenman, O. J., and Sterner, T. (1998). "What is the scope for environmental road pricing?" Road pricing, Traffic congestion, and the Environment, K. Button and E. T. Verhoef, eds., Edward Elgar Publishing, Massachusetts.

Sumalee, A. (2001). "Analysing the Design Criteria of Charging Cordons." *ITS working paper no 560*, Institute for Transport Studies, University of Leeds, Leeds.

Sumalee, A., Connors, R., and Watling, D. P. (2004). "Optimal toll design problem with improved behavioural equilibrium model: the case of Probit model." Mathematical and Computational Models for Congestion Charging, D. W. Hearn, S. Lawphongpanich, and M. J. Smith, eds.

Suresh, G., Vinod, V. V., and Sahu, S. (1995). "A genetic algorithm for facility layout." *International Journal of Production Research*, 33(12), 3411-3423.

Surry, P. D., and Radcliffe, N. J. (1997). "The COMOGA method: Constrained optimization by multipleobjective genetic algorithms." *Control and Cybernetics*, 26(3), 391-412.

Suwansirikul, C., Friesz, T. L., and Tobin, R. L. (1987). "Equilibrium decomposed optimization: a heuristic for the continuous network equilibrium design problem." *Transportation Science*, 21, 254-263.

TfL. (2003). "Congestion Charging: 6 months on." Transport for London, London.

TfL. (2004). "Congestion Charging: 12 months on." Transport for London, London.

Tobin, R. L. (1986). "Sensitivity analysis for variational inequalities." *Journal of Optimization Theory and Applications*, 48, 191-204.

Tobin, R. L., and Friesz, T. L. (1988). "Sensitivity Analysis for Equilibrium Network Flow." *Transportation Science*, 22(4), 242-250.

Transport, D. M. o. (1995). "Contours of implementation of congestion charging (Rekening Rijden)." Abstract of a Letter to Parliament from the Minister of Transport.

Transpotech. (1985). "Electronics Road Pricing Pilote Scheme prepared for the Hong Kong Government." Hong Kong.

Turner, D. (2001). "Centre piece." TOLLtrans, 24-26.

Tuy, H., Migdalas, A., and Varbrand, P. (1993). "A global optimization approach for the linear two-level programs." *Journal of Global Optimization*, 3(1-23).

Van Vliet, D. (1982). "SATURN - a modern assignment model." *Traffic Engineering and Control*, 23(12).

Verhoef, E. (2003a). "Inside the queue: hypercongestion and road pricing in a continuous time-continuous place model of traffic congestion." *Journal of Urban Economics*, 54, 531-564.

Verhoef, E. (2003b). "Speed-flow relations and cost functions for congested traffic: theory and empirical analyses." *TI 2003-064/3*, Tinbergen Institute, Amsterdam-Rotterdam.

Verhoef, E., Nijkamp, P., and Rietveld, P. (1995). "Second-Best Regulation of Road Transport Externalities." *Journal Of Transport Economics And Policy*, 29(N.2), 147-167.

Verhoef, E. T. (2002). "Second-best congestion pricing in general networks: Heuristic algorithms for finding second-best optimal toll levels and toll points." *Transportation Research Part B*, 36(8), 707-729.

Vicente, L. N., and Calamai, P. H. (1994). "Bilevel and multilevel programming: A bibliography review." *Journal of Global Optimization*, 3, 291-306.

Vicente, L. N., Savard, G., and Judice, J. J. (1994). "Descent Approaches for Quadratic Bilevel Programming." *Journal of Optimization Theory and Applications*, 81(2), 379-399.

Vickrey, W. S. (1963). "Pricing in urban and suburban transport." *American Economic Review: Papers and Proceeding*, 53(2), 452-465.

Vickrey, W. S. (1969). "Congestion theory and transport investment." *American Economic Review*, 59(2), 251-260.

Viegas, J. M. (2001). "Making urban road pricing acceptable and effective: searching for quality and equity in urban mobility." *Transport Policy*, 8(4), 289-294.

Vreeker, R., Nijkamp, P., and Ter Welle, C. (2002). "A multicriteria decision support methodology for evaluating airport expansion plans." *Transportation Research Part D: Transport and Environment*, 7(1), 27-47.

Walters, A. A. (1961). "The Theory and Measurement of Private and Social Cost of Highway Congestion." *Econometrica: Journal of the Econometric Society*, 29(4), 676-699.

Wardrop, J. (1952). "Some theoretical aspects of road traffic research." *Proc. of the Institute of Civil Engineers*, 1(2).

White, D. J., and Anandalingam, G. (1993). "A penalty function approach for solving bilevel linear program." *Journal of Global Optimization*, 3(397-419).

Whitley, D. (1989). "The GENITOR algorithm and selection pressure : Why rank based allocation of reproductive trial is best." Proceeding of the 3rd International Conference on Genetic Algorithms, J. Schaffer, ed., Morgan Kaufmann, Publishers, CA, USA.

Willoughby, C. (2001). "Singapore's motorization policies 1960-2000." *Transport Policy*, 8(2), 125-139.

Wilson, J. D. (1983). "Optimal road capacity in the presence of unpriced congestion." *Journal of Urban Economics*, 13, 337-357.

Xiong, Y., and Schneider, J. B. (1992). "Transportation Network Design Using a Cumulative Genetic Algorithm and Neural Network." *Transportation Research Record*, 44(9 Fig., 2 Tab., 12 Ref).

Yang, H. (1997). "Sensitivity Analysis for the Elastic-Demand Network Equilibrium Problem with Applications." *Transportation Research, Part B*, 70(25 Refs).

Yang, H., and Bell, M. G. H. (1997). "Traffic restraint, road pricing and network equilibrium." *Transportation Research Part B: Methodological*, 31(4), 303-314.

Yang, H., and Huang, H.-J. (1998). "Principle of marginal-cost pricing: how does it work in a general road network?" *Transportation Research Part A: Policy and Practice*, 32(1), 45-54.

Yang, H., and Zhang, X. (2002). "Multiclass Network Toll Design Problem with Social and Spatial Equity Constraints." *Journal of Transportation Engineering*, 128(5), p420-428.

Yang, H., Zhang, X., and Huang, H.-J. (2002). "Determination of optimal toll level and toll locations of alternative congestion pricing schemes." Transportation and Traffic Theory in the 21st Century: Proceeding of the 15th International Symposium on Transportation and Traffic Theory, M. A. P. Taylor, ed., Elsevier Science, New York.

Yin, Y. (2000). "Genetic-Algorithms-Based Approach for Bilevel Programming Models." *Journal of Transportation Engineering*, 126(2), p115-120.

Zhang, J. Z., and Liu, G. S. (2001). "A New Extreme Point Algorithm and Its Application in PSQP Algorithms for Solving Mathematical Programs with Linear Complementarity Constraints." *Journal of Global Optimization*, 19, 345-361.

Zhang, X., and Yang, H. (2004). "The optimal cordon-based network congestion pricing problem." *Transportation Research Part B*, 38(6), 517-537.

Zhou, G., Min, H., and Gen, M. (2002). "The balanced allocation of customers to multiple distribution centers in the supply chain network: a genetic algorithm approach." *Computers & Industrial Engineering*, 43(1-2), 251-261.

Zitzler, E., and Thiele, L. (1999). "Multiobjective Evolutionary Algorithms: A Comparative Case Study and the Strength Pareto Approach." *IEEE Transactions on Evolutionary Computation*, 3(4), 257-271.

APPENDIX A. QUESTIONAIRE ON JUDGMENTAL DESIGN CRITERIA FOR CHARGING CORDON

SECTION I

1. What are the objectives of using charging cordons in Leeds and what is the level of importance of each selected objective?

	High	Medium	Low	Not an objective
Reducing congestion	☐	☐	☐	☐
Environmental protection	☐	☐	☐	☐
City centre management	☐	☐	☐	☐
Increase efficiency	☐	☐	☐	☐
Redressing inequity in transport system	☐	☐	☐	☐
Other, please state	☐	☐	☐	☐

Answer the next question if you choose more than one objective in question 1. If not, go to question 3.

2. Are there any differences in the charging cordon to meet the different objectives of the scheme at the same time?

Yes ☐ No ☐

If the answer is yes, please explain the differences:

3. What is the general design of the proposed cordon of Leeds? Please give the detail in terms of:

Location of the cordon

Charge structure

4. Would you consider using the additional cordons or screenlines?

Yes ☐ No ☐

If yes, please explain the reasons, and define the additional cordons or screenlines. If no, please explain the reasons.

SECTION II

5. For each of the design aspects, do you agree that that design aspect should be included into the design of the charging cordon? If there is the other design aspects that should be considered, please state in the blank box.

Avoid the adverse impacts	Strongly Agree	Agree	Not consider	Disagree	Strongly Disagree
• The design should ensure the provision of sufficient alternative routes for drivers who want to bypass the charge area.	☐	☐	☐	☐	☐
• The design should avoid the dispersion of environmental or congestion problem to other areas.	☐	☐	☐	☐	☐
• The cordon should cover only the area having good public transport service.	☐	☐	☐	☐	☐
• The design should leave the facilities for interchange outside the cordon (e.g. park and ride or parking facility).	☐	☐	☐	☐	☐
• The design should ensure that all entry points to the charge area are charged or closed.	☐	☐	☐	☐	☐
• The design of the entry points should not be visually unattractive.	☐	☐	☐	☐	☐
• The design should place cordons at boundaries between land use types.	☐	☐	☐	☐	☐

Gain public acceptance	Strongly Agree	Agree	Not consider	Disagree	Strongly Disagree
• The cordon structure should be simple and easy to understand.	☐	☐	☐	☐	☐
• The charge structure should also be simple and easy to understand.	☐	☐	☐	☐	☐
• The charge should be at a level which is acceptable to the public.	☐	☐	☐	☐	☐
• The charge should be perceived as fair by the public.	☐	☐	☐	☐	☐
• The design should avoid the problem of local inequities (e.g. people just outside the cordon needing to access places just inside)	☐	☐	☐	☐	☐
• The design should avoid the problem of commercial inequities (e.g. with the same type of business, one is just inside the cordon and the other is just outside the cordon)	☐	☐	☐	☐	☐
• The design should aim at charging the traffic which contributes most to congestion and pollution.	☐	☐	☐	☐	☐
• The design should aim at charging the traffic which is of least benefit to the area.	☐	☐	☐	☐	☐
• The design should avoid charging the city's residents.	☐	☐	☐	☐	☐
• The design should avoid charging people from the low income area of the city.	☐	☐	☐	☐	☐

6. Based on the case of Leeds, are there any difficulties in the design to meet the required objectives and to satisfy the design issues mentioned above

Yes ☐ ☐ No

If the answer is Yes, please explain some difficulties in your city case.

SECTION III

7. How would your Local Authority spend the revenue generated from the charging scheme and why will you spend the revenues this way?

APPENDIX B. BACKGROUND OF THE CASE STUDY

In this section, the background information for each city is given. This information is based on the results of some parts of the in-depth interview which asked about the characteristics of each city.

Birmingham and West Midlands area

There are seven local authorities in the West Midland area including Birmingham, Coventry, Wolverhampton, Solihull, Dudley, Walsall, and Sandwell. All of these local authorities have a very strong collaboration in terms of transport planning in which there is a West Midlands committee as the political organisation providing an umbrella on the top of these seven local authorities. Each local authority has its own centre, but the whole area is a continuous conurbation.

Of these seven centres, Birmingham, Coventry, and Wolverhampton experience the most severe congestion problem in the city. It is agreed by the seven local authorities to look at the possibility of using congestion charging in the West Midlands area in order to raise revenue for improving public transport services and reducing congestion.

Bristol

Bristol is a city in the southern part of England. The city is surrounded by green-belt where the nearest centre is Cardiff. Therefore, the competition of the economy between the city centres is not so high. There are about 500,000 car movements everyday in and out the city centre alone and the traffic condition in the city centre is very congested and already at capacity. Bristol has two main ring roads, i.e. the inner ring road surrounding the core centre of the city and the outer ring road covering most of the city formed by the motorway network (M4 and M5 in the north and other trunk roads in the south of the city).

The major employment area is in the centre and north of the city. Congestion charging is one approach that the Council of Bristol is actively promoting as part of integrated transport measures to reduce demand for travel by and use of car. The council in conjunction with the local bus company ran a road user charging trial along the A4 Bath Road during 1998 in which this test aimed to examine both the technicalities of operating as well as how charging might change people's pattern of travel.

Durham

Durham has a fairly unique character compared to other cases. It is the country tows of a large county but has no strong commercial centre. There is a river running through the centre of the city forming a peninsula. The peninsula is considered as the historical area where there are the cathedral and castle, which are famous tourist places. The peninsula can only be accessed by a road which run through the market place. Beyond the market place there is no commercial activity.

The historical area on the peninsula has suffered from unnecessary traffic where a lot of traffic just drops people off at the market place area to go to banks or shops and then drives up to the peninsula area to wait and turn around to pick up people. The city council has decided to use road user charging in order to reduce the unnecessary trips in the historical area. The technological trial is on the progress and the system will be implemented in 2002.

Edinburgh

Edinburgh is the capital city of Scotland situated in the east of Scotland. The north of the city is bordered by the coast. There is a trunk road system generally surrounding the whole area of Edinburgh forming a very good outer ring road; there is also a tight inner ring road system surrounding the core of the city known as the Old and new towns which is considered as the historical area of the city. The city has two controlled parking zones including the inner parking zone which concentrates on the central area of the city and the outer parking zone which extends beyond the core of the city. The city has been considering the plan to implement road user charging scheme in order to reduce congestion in the city and raise the revenue to be used to improve public transport service.

Leeds

Leeds is one of the major cities situated in the north of England. Geographically, the city is quite separated from its neighbours, especially the city centre which is the major business and shopping area of the region. The inner ring road formed by a motorway system is regarded as the boundary of the central area. However, this ring road is still not completed yet; the completion of the ring road to the east of the city is on the plan. Similarly to the other big cities, Leeds faces the problem of increasing traffic and congestion in the city.

There are two types of congestion including the congestion on all radial routes, particularly on motorways, coming to the city centre in the peak period and the area inside the inner ring road which is congested almost all day. The University of Leeds is located just outside the inner ring but it is considered as a major trip attraction for inbound traffic in the peak period. The current problem is the lack of control in parking space in the city centre which causes failure to control the traffic demand in the city centre. The city has considered implementing a congestion charging scheme around the city centre in order to control the traffic demand and also reduce congestion in the city.

Manchester

There are nine districts in Greater Manchester district for which the geography of the area is considered as a polycentric conurbation area. Around Manchester, there are major important towns with their own shopping centres including Stockport and Bolton. This causes a very competitive situation for the city centre of Manchester with other centres and also the retail shopping area outside the city centre. Manchester city is a strange shape, long thin and rectangular. The city centre is only about two miles from the neighbouring local authority which is Salford. The worst congestion area is around the outer ring road whereas there are also some particular congestion areas in the cit centre. Manchester has joined the DTLR Congestion Charging Partnership but has not got a firm plan to implement a congestion charging scheme.

APPENDIX C. MATHEMATICAL NOTATIONS

Notations	Definitions
N	Set of nodes
A	Set of links
K	Set of O-D pairs
P	Set of paths
a	Link indices
i,j	Node indices
r	Origin node indices
s	Destination node indices
k	O-D pair indices
p	Path indices
d_k	Demand between O-D pair k
D_k	Demand function for O-D pair k
μ_k	Minimum travel cost between O-D pair k
v_a	Link flow
x_a^k	Multicommodity flow on link a from O-D pair k
$t_a(\mathrm{v})$	Travel time function for link a
$c_a(\mathrm{v})$	Travel cost function for link a
F_p	Path flow on path p
C_p	Path cost for path p
$\delta_{p,a}$	Dummy variable 1 if link a is on path p; 0 otherwise
$\Delta_{p,k}$	Dummy variable 1 if path p connects O-D pair k; 0 otherwise
b_i	$= 0$ if node i is neither origin nor destination node $= -d_k$ if node i is the origin of O-D pair k $= d_k$ if node i is the destination of O-D pair k
ψ_m	Objective function m
τ_a	Toll on link a
ε_a	Dummy variable 1 if link a is tolled; 0 otherwise
M	Number of objective functions

APPENDIX D. THE LINK COST AND OD DEMAND FUNCTIONS FOR THE EDINBURGH NETWORK

Link cost function

$$t_j = a + b \cdot \left(\frac{v_j}{c} \right)^n$$

Start node	End node	*a*	*b*	*c*	*n*	Start node	End node	*a*	*b*	*c*	*n*
100	200	203.14	63.45	2000	2.5	180	625	90.76	28.35	2000	2.5
100	515	500.09	105	4000	3	180	190	90.73	19.05	4000	3
100	101	27.86	11.57	1500	2	181	180	241.46	100.29	1500	2
101	100	27.86	11.57	1500	2	181	191	113.44	72	1500	1.5
101	102	167.17	69.43	1500	2	181	360	278.62	115.71	1500	2
101	201	238.23	151.2	1500	1.5	182	171	198.82	62.1	2000	2.5
102	101	167.17	69.43	1500	2	182	180	120.73	50.14	1500	2
102	110	92.87	38.57	1500	2	182	390	289.76	120.34	1500	2
102	103	34.03	21.6	1500	1.5	190	240	235.76	49.5	4000	3
103	102	34.03	21.6	1500	1.5	190	191	148.6	61.71	1500	2
103	111	272.26	172.8	1500	1.5	190	180	90.73	19.05	4000	3
110	102	92.87	38.57	1500	2	190	625	108.05	33.75	2000	2.5
110	111	176.45	73.29	1500	2	191	181	113.44	72	1500	1.5
110	303	124.79	79.2	1500	1.5	191	190	148.6	61.71	1500	2
111	103	272.26	172.8	1500	1.5	191	241	226.88	144	1500	1.5
111	110	176.45	73.29	1500	2	191	370	331.25	210.24	1500	1.5
111	112	125.38	52.07	1500	2	200	100	203.14	63.45	2000	2.5
112	111	125.38	52.07	1500	2	200	201	147.47	93.6	1500	1.5
112	155	32.5	13.5	1500	2	200	202	35.72	7.5	4000	3
112	113	164.49	104.4	1500	1.5	201	101	238.23	151.2	1500	1.5
113	112	164.49	104.4	1500	1.5	201	200	147.47	93.6	1500	1.5
113	114	192.85	122.4	1500	1.5	201	203	181.51	115.2	1500	1.5
114	113	192.85	122.4	1500	1.5	201	300	544.52	345.6	1500	1.5
114	115	102.1	64.8	1000	1.5	202	200	35.72	7.5	4000	3
114	121	119.11	75.6	1500	1.5	202	203	207.46	64.8	4000	3
115	114	102.1	64.8	1000	1.5	202	205	169.32	35.55	6000	3
115	120	97.51	40.5	1500	2	202	525	392.93	67.94	4000	3
115	155	78.94	32.79	1500	2	203	201	181.51	115.2	1500	1.5
120	115	97.51	40.5	1500	2	203	202	207.46	64.8	4000	3
120	121	27.86	11.57	1500	2	203	212	68.07	43.2	1500	1.5
120	124	45.38	28.8	1500	1.5	205	202	169.32	35.55	6000	3
121	114	119.11	75.6	1500	1.5	205	525	500.09	105	6000	3
121	120	27.86	11.57	1500	2	205	210	57.15	12	6000	3
121	122	83.59	34.71	1500	2	210	205	57.15	12	6000	3
122	121	83.59	34.71	1500	2	210	211	77.8	24.3	4000	3
122	123	73.74	46.8	1500	1.5	210	213	142.88	30	4000	3
122	140	328.76	136.54	2000	2.5	210	636	777.5	132.63	2000	2.5
123	122	73.74	46.8	1500	1.5	211	210	77.8	24.3	4000	3
123	124	34.03	21.6	1500	1.5	211	212	51.87	16.2	4000	3
123	125	181.51	115.2	1500	1.5	211	215	69.65	28.93	4000	3
123	330	266.59	169.2	1000	1.5	212	203	68.07	43.2	1500	1.5
124	120	45.38	28.8	1500	1.5	212	211	51.87	16.2	4000	3
124	123	34.03	21.6	1500	1.5	212	310	369.63	153.51	1500	2
124	145	198.52	126	1500	1.5	213	217	260.05	54.6	4000	3
125	123	181.51	115.2	1500	1.5	214	213	18.58	7.71	1500	2
125	325	102.1	64.8	1500	1.5	214	215	78.94	32.79	4000	3
125	126	28.36	18	4000	3	214	220	51.08	21.21	4000	3

126	125	28.36	18	4000	3	215	211	69.65	28.93	4000	3
126	145	102.1	64.8	1500	1.5	215	214	78.94	32.79	4000	3
126	320	45.38	28.8	4000	3	216	210	142.88	30	4000	3
140	122	328.76	136.54	2000	2.5	216	215	181.51	115.2	1500	1.5
140	151	351.67	223.2	1500	1.5	217	216	260.05	54.6	4000	3
140	141	50.15	20.83	2000	2.5	217	230	128.6	27	4000	3
140	330	191.72	121.68	1500	1.5	217	224	46.43	19.29	1500	2
141	140	50.15	20.83	2000	2.5	220	214	51.08	21.21	4000	3
141	150	148.6	61.71	2000	2.5	220	221	222.89	92.57	1500	2
141	380	124.79	79.2	1500	1.5	220	222	141.8	90	1500	1.5
145	124	198.52	126	1500	1.5	220	610	859.05	356.79	1500	2
145	126	102.1	64.8	1500	1.5	221	220	222.89	92.57	1500	2
145	303	351.67	223.2	1500	1.5	221	226	68.07	43.2	1500	1.5
145	155	141.8	90	1500	1.5	221	310	311.96	198	1500	1.5
150	141	148.6	61.71	2000	2.5	222	220	141.8	90	1500	1.5
150	151	130.02	54	1500	2	222	223	113.44	72	1500	1.5
150	160	89.3	18.75	4000	3	222	224	153.15	97.2	1500	1.5
150	380	80.34	33.36	1500	2	223	222	113.44	72	1500	1.5
151	140	351.67	223.2	1500	1.5	223	225	192.85	122.4	1500	1.5
151	150	130.02	54	1500	2	223	226	79.41	50.4	1500	1.5
151	152	74.29	30.86	1500	2	224	222	153.15	97.2	1500	1.5
152	151	74.29	30.86	1500	2	224	217	46.43	19.29	1500	2
152	153	96.43	61.2	1500	1.5	224	225	68.07	43.2	1500	1.5
152	161	34.03	21.6	1500	1.5	225	223	192.85	122.4	1500	1.5
153	152	96.43	61.2	1500	1.5	225	224	68.07	43.2	1500	1.5
153	162	62.39	39.6	1500	1.5	225	231	136.13	86.4	1500	1.5
153	172	250.75	104.14	1500	2	226	221	68.07	43.2	1500	1.5
153	174	436.49	181.29	1500	2	226	223	79.41	50.4	1500	1.5
155	112	32.5	13.5	1500	2	226	232	272.26	172.8	1000	1.5
155	115	78.94	32.79	1500	2	230	217	128.6	27	4000	3
155	145	141.8	90	1500	1.5	230	231	67.79	28.16	1500	2
160	150	89.3	18.75	4000	3	230	240	187.89	39.45	4000	3
160	161	176.45	73.29	1500	2	230	603	317.35	54.14	2000	2.5
160	170	192.18	40.35	4000	3	231	225	136.13	86.4	1500	1.5
160	390	122.52	77.76	1500	1.5	231	230	67.79	28.16	1500	2
161	152	34.03	21.6	1500	1.5	231	232	425.41	270	1500	1.5
161	160	176.45	73.29	1500	2	231	241	250.75	104.14	1500	2
161	162	68.07	43.2	1500	1.5	232	226	272.26	172.8	1000	1.5
162	153	62.39	39.6	1500	1.5	232	231	425.41	270	1500	1.5
162	161	68.07	43.2	1500	1.5	232	340	141.8	90	1500	1.5
162	171	170.16	108	1500	1.5	240	190	235.76	49.5	4000	3
170	160	192.18	40.35	4000	3	240	230	187.89	39.45	4000	3
170	171	34.58	10.8	2000	2.5	240	241	99.41	31.05	4000	3
170	173	178.6	37.5	4000	3	240	604	155.6	48.6	2000	2.5
170	180	251.47	52.8	4000	3	241	191	226.88	144	1500	1.5
171	162	170.16	108	1500	1.5	241	231	250.75	104.14	1500	2
171	170	34.58	10.8	2000	2.5	241	240	99.41	31.05	4000	3
171	182	198.82	62.1	2000	2.5	241	370	273.96	113.79	1500	2
172	153	250.75	104.14	1500	2	300	201	544.52	345.6	1500	1.5
172	173	79.33	13.53	2000	2.5	300	301	79.41	50.4	1500	1.5
172	621	397.63	124.2	2000	2.5	300	302	79.41	50.4	1500	1.5
173	170	178.6	37.5	4000	3	300	310	113.44	72	1500	1.5
173	172	79.33	13.53	2000	2.5	301	300	79.41	50.4	1500	1.5
173	174	214.33	45	4000	3	301	305	68.07	43.2	1500	1.5
174	153	436.49	181.29	1500	2	301	340	28.36	18	1500	1.5
174	173	214.33	45	4000	3	302	300	79.41	50.4	1500	1.5
174	620	857.3	180	4000	3	302	303	277.93	176.4	1500	1.5
180	170	251.47	52.8	4000	3	302	305	11.34	7.2	4000	3
180	181	241.46	100.29	1500	2	303	110	124.79	79.2	1500	1.5
180	182	120.73	50.14	1500	2	303	145	351.67	223.2	1500	1.5

180	621	139.3	57.86	1500	2	303	302	277.93	176.4	1500	1.5
321	350	187.18	118.8	1500	1.5	305	301	68.07	43.2	1500	1.5
325	125	102.1	64.8	1500	1.5	305	302	11.34	7.2	4000	3
325	320	158.82	100.8	1500	1.5	305	321	130.46	82.8	4000	3
325	330	147.47	93.6	1500	1.5	310	212	369.63	153.51	1500	2
330	123	266.59	169.2	1000	1.5	310	221	311.96	198	1500	1.5
330	140	191.72	121.68	1500	1.5	310	300	113.44	72	1500	1.5
330	325	147.47	93.6	1500	1.5	310	340	158.82	100.8	1500	1.5
330	380	305.16	193.68	1500	1.5	320	126	45.38	28.8	4000	3
340	232	141.8	90	1500	1.5	320	321	5.67	3.6	4000	3
340	301	28.36	18	1500	1.5	320	325	158.82	100.8	1500	1.5
340	310	158.82	100.8	1500	1.5	321	305	130.46	82.8	4000	3
340	350	215.54	136.8	1500	1.5	321	320	5.67	3.6	4000	3
350	321	187.18	118.8	1500	1.5	602	601	352.91	146.57	1500	2
350	340	215.54	136.8	1500	1.5	602	623	785.43	133.98	2000	2.5
350	360	217.81	138.24	1500	1.5	603	230	317.35	54.14	2000	2.5
350	371	181.51	115.2	1500	1.5	603	601	388.99	121.5	2000	2.5
360	181	278.62	115.71	1500	2	603	604	111.07	18.95	2000	2.5
360	350	217.81	138.24	1500	1.5	604	240	155.6	48.6	2000	2.5
360	371	45.38	28.8	1500	1.5	604	603	111.07	18.95	2000	2.5
360	390	359.61	228.24	1500	1.5	604	624	482.93	200.57	1500	2
370	191	331.25	210.24	1500	1.5	605	525	642.98	135	4000	3
370	241	273.96	113.79	1500	2	605	633	213.61	88.71	4000	3
370	371	63.53	40.32	1500	1.5	605	637	396.5	83.25	6000	3
371	350	181.51	115.2	1500	1.5	610	220	859.05	356.79	1500	2
371	360	45.38	28.8	1500	1.5	610	636	329.69	136.93	1500	2
371	370	63.53	40.32	1500	1.5	620	174	857.3	180	4000	3
380	141	124.79	79.2	1500	1.5	620	621	1713.7	292.33	2000	2.5
380	150	80.34	33.36	1500	2	621	172	397.63	124.2	2000	2.5
380	330	305.16	193.68	1500	1.5	621	180	139.3	57.86	1500	2
380	390	106.8	44.36	1500	2	621	620	1713.7	292.33	2000	2.5
390	160	122.52	77.76	1500	1.5	621	622	102.1	64.8	1500	1.5
390	182	289.76	120.34	1500	2	622	621	102.1	64.8	1500	1.5
390	360	359.61	228.24	1500	1.5	622	623	96.43	61.2	1500	1.5
390	380	106.8	44.36	1500	2	622	624	78.94	32.79	1500	2
505	510	339.35	71.25	6000	3	623	602	785.43	133.98	2000	2.5
505	511	465.11	295.2	1500	1.5	623	622	96.43	61.2	1500	1.5
505	600	1264.5	218.65	4000	3	623	624	60.51	18.9	2000	2.5
510	505	339.35	71.25	6000	3	624	604	482.93	200.57	1500	2
510	511	468.99	194.79	4000	3	624	622	78.94	32.79	1500	2
510	512	67.87	14.25	4000	3.1	624	623	60.51	18.9	2000	2.5
511	505	465.11	295.2	1500	1.5	624	625	73.48	22.95	2000	2.5
511	510	468.99	194.79	4000	3	625	180	90.76	28.35	2000	2.5
512	510	67.87	14.25	4000	3.1	625	190	108.05	33.75	2000	2.5
512	515	505.68	157.95	4000	3	625	624	73.48	22.95	2000	2.5
512	600	2161	675	2000	2.5	630	515	595.02	101.5	2000	2.5
515	100	500.09	105	4000	3	630	520	457.23	96	6000	3
515	512	505.68	157.95	4000	3	630	631	1586.7	270.68	2000	2.5
515	520	325.28	55.49	2000	2.5	631	630	1586.7	270.68	2000	2.5
515	630	595.02	101.5	2000	2.5	631	632	476.02	81.2	2000	2.5
520	515	325.28	55.49	2000	2.5	631	637	709.01	450	1500	1.5
520	525	157.17	33	6000	3	632	631	476.02	81.2	2000	2.5
520	630	457.23	96	6000	3	632	633	896.2	372.21	1500	2
525	202	392.93	67.94	4000	3	632	634	631.51	262.29	1500	2
525	205	500.09	105	6000	3	632	637	282.19	59.25	6000	3
525	520	157.17	33	6000	3	633	605	213.61	88.71	4000	3
525	605	642.98	135	4000	3	633	632	896.2	372.21	1500	2
600	505	1264.5	218.65	4000	3	633	635	204.31	84.86	4000	3
600	512	2161	675	2000	2.5	634	632	631.51	262.29	1500	2
601	602	352.91	146.57	1500	2	634	635	1160.9	482.14	1500	2

601	603	388.99	121.5	2000	2.5	635	634	1160.9	482.14	1500	2
636	635	315.76	131.14	1500	2	635	636	315.76	131.14	1500	2
637	605	396.5	83.25	6000	3	635	633	204.31	84.86	4000	3
637	631	709.01	450	1500	1.5	636	210	777.5	132.63	2000	2.5
637	632	282.19	59.25	6000	3	636	610	329.69	136.93	1500	2

OD demand function

$$d_k = d_k^0 \cdot \left(\frac{\mu_k}{\mu_k^0} \right)^{\beta}$$

Origin	Destination	d_k^0	μ_k^0	β	Origin	Destination	d_k^0	μ_k^0	β
321	340	608.15	414	-0.58	121	310	1244.5	642	-0.58
321	121	1000.7	943	-0.58	121	303	3665.7	142	-0.58
321	140	795.13	391	-0.58	121	633	2717.4	248	-0.58
321	181	1329.3	180	-0.58	121	525	2839.3	98	-0.58
321	371	951.93	253	-0.58	121	601	1906.5	175	-0.58
321	231	1470.9	226	-0.58	121	621	1323.4	67	-0.58
321	222	1490.6	132	-0.58	121	153	2887.7	250	-0.58
321	212	1634.5	325	-0.58	121	620	2322.4	329	-0.58
321	201	1512.3	161	-0.58	121	515	2921.2	179	-0.58
321	103	774.97	158	-0.58	121	630	4141	167	-0.58
321	113	606.66	947	-0.58	140	321	1335.3	834	-0.58
321	350	786.47	179	-0.58	140	340	1943.4	274	-0.58
321	310	1006.5	560	-0.58	140	121	1221.1	732	-0.58
321	303	3161.9	116	-0.58	140	181	1330.5	274	-0.58
321	633	2213.6	230	-0.58	140	371	1313.7	270	-0.58
321	525	2321.9	56	-0.58	140	231	1853.9	230	-0.58
321	601	1791.9	178	-0.58	140	222	2253.3	82	-0.58
321	621	1341.2	49	-0.58	140	212	2884.1	353	-0.58
321	153	2902	183	-0.58	140	201	2847.6	144	-0.58
321	620	2443.4	239	-0.58	140	103	1589.3	100	-0.58
321	515	2835.4	136	-0.58	140	113	1941.9	697	-0.58
321	630	3637.1	130	-0.58	140	350	2121.7	187	-0.58
340	321	970.07	450	-0.58	140	310	2006.9	321	-0.58
340	121	1970.7	336	-0.58	140	303	4033.3	115	-0.58
340	140	1765.2	150	-0.58	140	633	3261.8	232	-0.58
340	181	1598.9	123	-0.58	140	525	2114.5	236	-0.58
340	371	1221.6	155	-0.58	140	601	1181.7	208	-0.58
340	231	862.76	318	-0.58	140	621	598.63	21	-0.58
340	222	882.43	93	-0.58	140	153	2163	82	-0.58
340	212	1132.5	192	-0.58	140	620	3379.1	109	-0.58
340	201	1194.8	124	-0.58	140	515	3883.6	58	-0.58
340	103	1597.2	87	-0.58	140	630	4685.3	57	-0.58
340	113	876.31	576	-0.58	181	321	2046.7	423	-0.58
340	350	284.52	234	-0.58	181	340	1954.5	205	-0.58
340	310	1098.8	267	-0.58	181	121	2730.4	510	-0.58
340	303	2667.6	52	-0.58	181	140	1509.3	280	-0.58
340	633	1896.1	102	-0.58	181	371	696.46	343	-0.58
340	525	1713.7	22	-0.58	181	231	908.73	253	-0.58
340	601	2066.7	55	-0.58	181	222	1517.1	96	-0.58
340	621	2236.3	21	-0.58	181	212	2149.9	291	-0.58
340	153	3712.7	96	-0.58	181	201	2560.3	123	-0.58
340	620	2188.4	133	-0.58	181	103	2821.6	74	-0.58
340	515	2517.9	55	-0.58	181	113	1324.7	547	-0.58
340	630	3319.6	52	-0.58	181	350	2239.1	163	-0.58
121	321	1107.8	1227	-0.58	181	310	3053.2	252	-0.58

121	340	1715.9	408	-0.58	181	303	3299.1	133	-0.58
121	140	724.74	443	-0.58	181	633	2527.6	338	-0.58
121	181	2055.2	275	-0.58	181	525	1380.3	38	-0.58
121	371	2038.4	434	-0.58	181	601	529.7	62	-0.58
121	231	2578.7	361	-0.58	181	621	963.43	35	-0.58
121	222	2598.4	188	-0.58	181	153	2252.1	81	-0.58
121	212	2742.3	500	-0.58	181	620	3027.7	102	-0.58
121	201	2208.1	271	-0.58	181	515	3149.4	93	-0.58
121	103	368.18	189	-0.58	181	630	3951.1	84	-0.58
121	113	1714.5	830	-0.58	371	321	1447.2	462	-0.58
121	350	1894.3	309	-0.58	371	340	1355.1	230	-0.58
371	222	1506.7	112	-0.58	371	121	2447.8	540	-0.58
371	212	2139.5	289	-0.58	371	140	1226.7	225	-0.58
371	201	2549.9	157	-0.58	371	181	432.32	238	-0.58
371	103	2222.2	80	-0.58	371	231	898.35	331	-0.58
371	113	725.26	793	-0.58	310	103	498.36	188	-0.58
371	350	1639.6	188	-0.58	310	113	1822.7	477	-0.58
371	310	2453.8	326	-0.58	310	350	1075.8	248	-0.58
371	303	3288.7	95	-0.58	310	303	2951	108	-0.58
371	633	2517.2	189	-0.58	310	633	2002.7	148	-0.58
371	525	1370	56	-0.58	310	525	2660.1	21	-0.58
371	601	962.04	80	-0.58	310	601	2674.9	112	-0.58
371	621	1069.7	30	-0.58	310	621	2224.2	58	-0.58
371	153	2546.2	88	-0.58	310	153	3785	176	-0.58
371	620	3017.3	106	-0.58	310	620	1607.6	226	-0.58
371	515	3139	87	-0.58	310	515	2206.4	161	-0.58
371	630	3940.7	73	-0.58	310	630	3426.2	152	-0.58
231	321	2402.9	540	-0.58	303	321	4892.3	681	-0.58
231	340	1432.8	396	-0.58	303	340	4048.9	255	-0.58
231	121	3403.6	495	-0.58	303	121	5325.4	529	-0.58
231	140	2183.5	197	-0.58	303	140	5376.3	161	-0.58
231	181	914.71	199	-0.58	303	181	4280.6	152	-0.58
231	371	1319.9	327	-0.58	303	371	4685.8	151	-0.58
231	222	608.39	189	-0.58	303	231	3435.6	307	-0.58
231	212	1241.1	468	-0.58	303	222	2786.5	99	-0.58
231	201	1651.5	203	-0.58	303	212	2391.7	962	-0.58
231	103	3030	100	-0.58	303	201	2367.5	294	-0.58
231	113	2045.2	591	-0.58	303	103	3957.8	165	-0.58
231	350	1540.9	285	-0.58	303	113	4925.2	390	-0.58
231	310	2531.6	325	-0.58	303	350	3624.2	262	-0.58
231	303	2422.2	139	-0.58	303	310	4508.3	427	-0.58
231	633	1650.7	149	-0.58	303	633	1373.6	91	-0.58
231	525	850.9	20	-0.58	303	525	4043.7	49	-0.58
231	601	1203.9	58	-0.58	303	601	4396.7	75	-0.58
231	621	1637.7	51	-0.58	303	621	4830.4	6	-0.58
231	153	2926.3	136	-0.58	303	153	6119.1	12	-0.58
231	620	2150.8	168	-0.58	303	620	1873.7	13	-0.58
231	515	2272.4	145	-0.58	303	515	1995.4	255	-0.58
231	630	3074.2	132	-0.58	303	630	903.36	208	-0.58
222	321	2471.9	343	-0.58	633	321	3518.7	201	-0.58
222	340	1501.8	258	-0.58	633	340	2675.4	91	-0.58
222	121	3472.5	377	-0.58	633	121	3951.8	193	-0.58
222	140	2685.5	94	-0.58	633	140	4002.7	34	-0.58
222	181	1564	130	-0.58	633	181	2907	23	-0.58
222	371	1969.2	171	-0.58	633	371	3312.2	44	-0.58
222	231	649.28	236	-0.58	633	231	2062	75	-0.58
222	212	632.74	586	-0.58	633	222	1413	57	-0.58
222	201	1043.1	146	-0.58	633	212	1018.1	266	-0.58
222	103	2818.8	93	-0.58	633	201	993.85	238	-0.58
222	113	2378.1	314	-0.58	633	103	2584.2	144	-0.58

222	350	1252.6	248	-0.58	633	113	3551.8	128	-0.58
222	310	2600.5	229	-0.58	633	350	2250.7	63	-0.58
222	303	1990.5	109	-0.58	633	310	3134.6	154	-0.58
222	633	1219	49	-0.58	633	303	948.26	36	-0.58
222	525	1352.9	8	-0.58	633	525	2670.1	12	-0.58
222	601	1706	31	-0.58	633	601	3023.1	30	-0.58
222	621	2139.7	27	-0.58	633	621	3456.8	6	-0.58
222	153	3428.3	55	-0.58	633	153	4745.5	22	-0.58
222	620	1719.1	72	-0.58	633	620	500.1	30	-0.58
222	515	1840.8	74	-0.58	633	515	621.78	76	-0.58
222	630	2642.5	65	-0.58	633	630	1423.5	66	-0.58
212	321	2500.8	292	-0.58	525	321	3442.9	430	-0.58
212	340	1657.4	215	-0.58	525	340	2620.8	165	-0.58
212	121	3501.4	368	-0.58	525	121	3862.7	365	-0.58
212	140	3237.3	169	-0.58	525	140	2641.6	143	-0.58
212	181	2136.2	136	-0.58	525	181	1590.6	137	-0.58
212	371	2541.4	173	-0.58	525	371	1995.8	199	-0.58
212	231	1221.5	265	-0.58	525	231	1187.9	162	-0.58
212	222	572.24	295	-0.58	525	222	1587.3	95	-0.58
212	201	410.4	404	-0.58	525	212	2218.1	327	-0.58
212	103	2186.1	138	-0.58	525	201	2628.5	154	-0.58
212	113	2533.7	237	-0.58	525	103	4217.9	111	-0.58
212	350	1232.7	314	-0.58	525	113	2721	360	-0.58
212	310	2580.6	281	-0.58	525	350	2728.8	162	-0.58
212	303	1535.1	141	-0.58	525	310	3719.6	241	-0.58
212	633	763.59	164	-0.58	525	303	3367.3	16	-0.58
212	525	1904.7	42	-0.58	525	633	2595.8	58	-0.58
212	601	2257.7	59	-0.58	525	601	1390.2	12	-0.58
212	621	2691.4	91	-0.58	525	621	2094.5	47	-0.58
212	153	3980.1	357	-0.58	525	153	3112.6	85	-0.58
212	620	1263.7	468	-0.58	525	620	3095.9	106	-0.58
212	515	1385.4	265	-0.58	525	515	3217.6	113	-0.58
212	630	2187.1	246	-0.58	525	630	4019.3	109	-0.58
201	321	2524.9	340	-0.58	601	321	2798.8	825	-0.58
201	340	1800.1	187	-0.58	601	340	2706.7	258	-0.58
201	121	3143.3	331	-0.58	601	121	2792.1	634	-0.58
201	140	3320	133	-0.58	601	140	1571	434	-0.58
201	181	2670.4	90	-0.58	601	181	752.12	385	-0.58
201	371	3021.7	139	-0.58	601	371	1351.6	415	-0.58
201	231	1755.7	178	-0.58	601	231	1275.5	215	-0.58
201	222	1106.4	124	-0.58	601	222	1674.9	54	-0.58
201	212	534.18	628	-0.58	601	212	2305.7	262	-0.58
201	103	1775.7	178	-0.58	601	201	2716.1	193	-0.58
201	113	2676.4	328	-0.58	601	103	3159.3	93	-0.58
201	350	1766.8	224	-0.58	601	113	2076.8	585	-0.58
201	310	2326.2	320	-0.58	601	350	2816.4	204	-0.58
201	303	1649.6	118	-0.58	601	310	3470.4	313	-0.58
201	633	701.3	111	-0.58	601	303	3454.9	43	-0.58
201	525	2438.8	25	-0.58	601	633	2683.4	341	-0.58
201	601	2791.9	70	-0.58	601	525	1308.1	39	-0.58
201	621	3225.6	58	-0.58	601	621	704.26	48	-0.58
201	153	4514.3	130	-0.58	601	153	1722.4	58	-0.58
201	620	993.64	171	-0.58	601	620	3183.5	77	-0.58
201	515	1323.1	166	-0.58	601	515	3305.2	119	-0.58
201	630	2124.8	154	-0.58	601	630	4106.9	111	-0.58
103	321	1063.3	363	-0.58	621	321	2140.3	250	-0.58
103	340	1671.4	152	-0.58	621	340	2611.9	77	-0.58
103	121	405.29	272	-0.58	621	121	2087.8	304	-0.58
103	140	1130	85	-0.58	621	140	866.69	523	-0.58
103	181	2392.5	65	-0.58	621	181	1019.6	98	-0.58

103	371	2015.2	89	-0.58	621	371	1353.8	115	-0.58
103	231	2534.2	100	-0.58	621	231	1543	56	-0.58
103	222	2159.4	91	-0.58	621	222	1942.4	8	-0.58
103	212	1703.8	268	-0.58	621	212	2573.2	97	-0.58
103	201	1169.6	201	-0.58	621	201	2983.6	44	-0.58
103	113	1669.9	114	-0.58	621	103	2456	46	-0.58
103	350	1800.9	111	-0.58	621	113	1982.1	142	-0.58
103	310	725.04	281	-0.58	621	350	2896.4	53	-0.58
103	303	2627.2	66	-0.58	621	310	2812	106	-0.58
103	633	1678.9	80	-0.58	621	303	3722.4	48	-0.58
103	525	3231.4	25	-0.58	621	633	2950.9	69	-0.58
103	601	2298.6	33	-0.58	621	525	1803.7	366	-0.58
103	621	1728.7	31	-0.58	621	601	648.56	59	-0.58
103	153	3293	104	-0.58	621	153	1564.4	24	-0.58
103	620	1283.8	131	-0.58	621	620	3451	33	-0.58
103	515	1882.7	83	-0.58	621	515	3572.7	20	-0.58
103	630	3102.4	78	-0.58	621	630	4374.4	14	-0.58
113	321	721.93	766	-0.58	153	321	3502.8	733	-0.58
113	340	629.83	374	-0.58	153	340	3952.8	215	-0.58
113	121	1722.6	566	-0.58	153	121	3496	384	-0.58
113	140	1517	299	-0.58	153	140	2274.9	325	-0.58
113	181	722.61	289	-0.58	153	181	2105.2	110	-0.58
113	371	345.27	471	-0.58	153	371	2694.7	168	-0.58
113	231	1243.6	309	-0.58	153	231	2628.6	84	-0.58
113	222	1512.3	104	-0.58	153	222	3028	46	-0.58
113	212	1762.4	224	-0.58	153	212	3658.8	123	-0.58
113	201	1824.6	164	-0.58	153	201	4069.2	76	-0.58
113	103	1496.9	56	-0.58	153	103	3863.5	67	-0.58
113	350	914.34	152	-0.58	153	113	3323	478	-0.58
113	310	1728.5	246	-0.58	153	350	4169.5	66	-0.58
113	303	3297.4	102	-0.58	153	310	4174.5	205	-0.58
113	633	2525.9	171	-0.58	153	303	4808	20	-0.58
113	525	1715.2	42	-0.58	153	633	4036.5	19	-0.58
113	601	1252.3	122	-0.58	153	525	2889.2	93	-0.58
113	621	1360	26	-0.58	153	621	1409.9	29	-0.58
113	153	2836.4	119	-0.58	153	620	4536.6	60	-0.58
113	620	2818.3	148	-0.58	153	515	4658.3	83	-0.58
113	515	3147.7	77	-0.58	153	630	5460	71	-0.58
113	630	3949.4	69	-0.58	620	321	4027.4	107	-0.58
350	321	1268.1	334	-0.58	620	340	3302.7	32	-0.58
350	340	424.72	254	-0.58	620	121	3766.4	70	-0.58
350	121	2268.8	348	-0.58	620	140	4392.2	23	-0.58
350	140	2063.2	123	-0.58	620	181	3677.5	35	-0.58
350	181	2023.6	93	-0.58	620	371	4082.7	26	-0.58
350	371	1646.3	166	-0.58	620	231	2832.5	46	-0.58
350	231	1114.6	224	-0.58	620	222	2183.6	29	-0.58
350	222	773.42	136	-0.58	620	212	1728	106	-0.58
350	212	848.01	419	-0.58	620	201	1502.6	68	-0.58
350	201	1258.4	199	-0.58	620	103	2398.9	32	-0.58
350	103	1846.3	81	-0.58	620	113	4179	55	-0.58
350	113	1301	279	-0.58	620	350	2960.7	36	-0.58
350	310	1347.9	285	-0.58	620	310	2949.3	82	-0.58
350	303	2383.1	62	-0.58	620	303	1718.8	43	-0.58
350	633	1611.6	84	-0.58	620	633	770.57	48	-0.58
350	525	1965.5	25	-0.58	620	525	3440.7	7	-0.58
350	601	2318.6	44	-0.58	620	601	3793.7	22	-0.58
350	621	2609.3	34	-0.58	620	621	4227.4	1	-0.58
350	153	4041	116	-0.58	620	153	5516.1	2	-0.58
350	620	2111.7	151	-0.58	620	515	598.83	113	-0.58
350	515	2233.4	87	-0.58	620	630	2194.1	54	-0.58

350	630	3035.1	86	-0.58	515	321	4186.9	385	-0.58
310	321	1216	753	-0.58	515	340	3343.6	142	-0.58
310	340	946.39	314	-0.58	515	121	4367.7	303	-0.58
310	121	1346.9	617	-0.58	515	140	4670.8	87	-0.58
310	140	1678.2	190	-0.58	515	181	3575.1	91	-0.58
310	181	2545.3	123	-0.58	515	371	3980.3	86	-0.58
310	371	2168	234	-0.58	515	231	2730.1	173	-0.58
310	231	1809.2	286	-0.58	515	222	2081.2	56	-0.58
310	222	1828.8	134	-0.58	515	212	1686.3	544	-0.58
310	212	1923.8	391	-0.58	515	201	1662	166	-0.58
310	201	1493.4	253	-0.58	515	103	3000.1	98	-0.58
632	140	7725.5	62	-0.58	515	113	4219.9	221	-0.58
632	181	7010.9	86	-0.58	515	350	2918.9	146	-0.58
632	371	7416.1	76	-0.58	515	310	3550.6	245	-0.58
632	231	6165.9	127	-0.58	515	303	1616.4	89	-0.58
632	222	5517	80	-0.58	515	633	668.16	52	-0.58
632	212	5061.4	312	-0.58	515	525	3338.2	23	-0.58
632	201	4836	187	-0.58	515	601	3691.3	48	-0.58
632	103	5732.3	82	-0.58	515	621	4125	1	-0.58
632	113	7512.4	154	-0.58	515	153	5413.7	5	-0.58
632	350	6294.1	100	-0.58	515	620	601.23	8	-0.58
632	310	6282.7	221	-0.58	515	630	2091.7	1337	-0.58
632	303	5052.2	104	-0.58	630	321	5389.5	535	-0.58
632	633	4103.9	112	-0.58	630	340	4546.2	208	-0.58
632	525	6774	14	-0.58	630	121	5822.5	426	-0.58
632	601	7127.1	64	-0.58	630	140	5873.4	129	-0.58
632	621	7560.8	150	-0.58	630	181	4777.7	126	-0.58
632	153	8849.5	1451	-0.58	630	371	5182.9	125	-0.58
632	620	3333.4	556	-0.58	630	231	3932.7	241	-0.58
632	515	3932.2	2	-0.58	630	222	3283.8	80	-0.58
632	630	5527.5	2	-0.58	630	212	2888.9	777	-0.58
511	321	8532.2	289	-0.58	630	201	2864.6	235	-0.58
511	340	7807.5	89	-0.58	630	103	4454.9	137	-0.58
511	121	8271.2	185	-0.58	630	113	5422.5	309	-0.58
511	140	8896.9	60	-0.58	630	350	4121.5	209	-0.58
511	181	8182.3	78	-0.58	630	310	5005.4	341	-0.58
511	371	8587.5	71	-0.58	630	303	898.62	121	-0.58
511	231	7337.3	122	-0.58	630	633	1870.8	80	-0.58
511	222	6688.4	75	-0.58	630	525	4540.8	39	-0.58
511	212	6232.8	303	-0.58	630	601	4893.9	67	-0.58
511	201	6007.4	178	-0.58	630	621	5327.6	3	-0.58
511	103	6903.6	79	-0.58	630	153	6616.3	10	-0.58
511	113	8683.8	147	-0.58	630	620	2370.9	13	-0.58
511	350	7465.4	101	-0.58	630	515	2082.3	630	-0.58
511	310	7454.1	215	-0.58	632	321	7360.8	299	-0.58
511	303	6223.6	105	-0.58	632	340	6636.1	92	-0.58
511	633	5275.3	114	-0.58	632	121	7099.8	193	-0.58
511	525	7945.4	12	-0.58	511	153	10021	1333	-0.58
511	601	8298.5	61	-0.58	511	620	4504.8	832	-0.58
511	621	8732.2	86	-0.58	511	630	6698.8	2	-0.58
					321	340	608.15	414	-0.58