Computer-Aided Civil and Infrastructure Engineering **22** (2007) 31–43

A Genetic Algorithm Approach for Optimizing Traffic Control Signals Considering Routing

Fitsum Teklu

Institute for Transport Studies, University of Leeds, 38 University Road, Leeds, LS2 9JT, United Kingdom

Agachai Sumalee†

Department of Civil Engineering, The Hong Kong Polytechnic University, Hunghom, Hong Kong

&

David Watling*

Institute for Transport Studies, University of Leeds, 38 University Road, Leeds, LS2 9JT, United Kingdom

Abstract: *It is well known that coordinated, area-wide traffic signal control provides great potential for improvements in delays, safety, and environmental measures. However, an aspect of this problem that is commonly neglected in practice is the potentially confounding effect of drivers re-routing in response to changes in travel times on competing routes, brought about by the changes to the signal timings. This article considers the problem of optimizing signal green and cycle timings over an urban network, in such a way that the optimization anticipates the impact on traffic routing patterns. This is achieved by including a network equilibrium model as a constraint to the optimization. A Genetic Algorithm (GA) is devised for solving the resulting problem, using total travel time across the network as an illustrative fitness function, and with a widely used traffic simulation-assignment model providing the equilibrium flows. The procedure is applied to a case study of the city of Chester in the UK, and the performance of the algorithms is analyzed with respect to the parameters of the GA method. The results show a better performance of the signal timing as optimized by the GA method as compared to a method that does not consider rerouting. This improvement is found to be more significant with a more congested network whereas under a relatively mild congestion situation the improvement is not very clear.*

†The work was carried out while the author was at the University of Leeds.
*To whom correspondence should be addressed. E-mail: *d.p.watling@its.leeds.ac.uk*.

1 INTRODUCTION

In urban road networks, traffic signals have been used to control vehicle movements so as to reduce congestion, improve safety, and enable specific strategies such as minimizing delays, prioritizing public transport, and improving environmental pollution (IHT, 1997). Through the years, procedures for determining optimum signal timings have been developed and continuously improved. Early methods such as that of Webster (1958) only considered a single signalized junction in isolation. Later, fixed time strategies were developed that optimized a group of signalized junctions using historical flow data (e.g., TRANSYT: Robertson, 1969). In some cities, real time traffic flow data has also been used for optimization in methods commonly referred to as demand-responsive strategies (e.g., SCOOT: Hunt et al., 1981). The focus of this research is on fixed time signal plans.

As mentioned earlier, fixed time plans use historical flows observed on links (through traffic counts) to optimize signal timings. One shortcoming of such optimization procedures is the assumption that the flows for which the optimal timings are calculated will remain unchanged after the new timings are implemented. This may not be a valid assumption if the implementation of such timings substantially improves the journey time on a certain route, since the users of alternative routes may divert to the improved route as a result. Such effects have been observed as a consequence of area wide traffic control schemes (e.g., Almond and Lott, 1968).

From a modeling viewpoint, the impact of signal time changes on changes in link flows could also be explained from the perspective of traffic assignment theory. In a transportation network, where traffic signals are explicitly modeled, drivers' route choices in a network could be assumed to follow a Wardrop user equilibrium (UE) condition (Wardrop, 1952). It then follows that a change in traffic signal timings disturbs the equilibrium by altering the route costs, hence changing the route flows. Such an equilibrium model may thus be used as a part of the signal optimization procedure (formally, as a constraint to the problem), to reflect the anticipated impact of changes to the timings on the flow patterns—that is, the flows are determined endogenously, inside the optimization process.

The problem of determining optimum signal timings while anticipating the equilibrium response of drivers is an instance of the network design problem (NDP). An NDP is concerned with improving an existing network so that some total network performance measure is optimized with respect to some discrete or continuous design variables, while considering users' response to the improvement. Some examples of NDPs include the capacity enhancement problem (Abdulaal and LeBlanc, 1979), the signal-setting problem (Chiou, 1999; Ceylan and Bell, 2004), toll pricing (Shepherd and Sumalee, 2004), and the charging cordon design problem (Sumalee, 2004).

NDPs are characterized by the so-called bi-level structure. On the upper level, a transport planner is assumed to "design the network," by selecting values for the design parameters so as to optimize some measure of total social cost (e.g., total delay, pollution, social welfare). Road users respond to that design in the lower level by altering their travel choices so as to minimize their own travel costs, which may not agree with the planner's view of the most appropriate costs.[1] Formally, the planner's optimization in the upper level is constrained by the lower level equilibrium problem. Such problems are known to be one of the most challenging mathematical problems in the optimization field, due to the non-smoothness of the objective function (not differentiable everywhere), coupled with the non-convexity of the feasible region. These properties imply the potential existence of multiple optima, as well as great difficulty in devising robust and efficient methods for even finding local optima (see, Luo et al., 1996 for more details).

A wide range of solution methods have been applied in an attempt to devise an efficient technique to solve NDPs, ranging over heuristic iterative methods (Steenbrink, 1974; Allsop, 1974; Suwansirikul et al., 1987), linearization methods (LeBlanc and Boyce, 1986; Ben-Ayed et al., 1988), sensitivity-based methods (Friesz et al., 1990; Yang, 1997), Karush–Kuhn–Tucker based methods (Marcotte, 1986; Verhoef, 2002), methods developed from the system optimal solution (Dantzig et al., 1979; Marcotte, 1981; Bergendorff et al., 1997; Hearn and Ramana, 1998), marginal function method (Meng et al., 2001), cutting plane method (Lawphongpanich and Hearn, 2004), and stochastic search methods (Friesz et al., 1992; Cree et al., 1998; Ceylan and Bell, 2004).

This article presents a solution method for solving a signal timing based NDP using a meta-heuristic optimization method, namely a Genetic Algorithm (GA). GA is an evolutionary optimization technique, formulated on the basis of the mechanics of natural selection and evolution. It offers great flexibility in solving such optimization problems as it does not require any information on the gradient of the objective function and has the ability to move out of local optima. An additional motivation for using GA for this problem is the simulation-based framework (as opposed to an analytic one) commonly employed in signal setting methods (e.g., TRANSYT and SATURN) that do not use explicit mathematical relations, which negates the use of derivative-based optimization methods. GA has been used widely in the transportation field, for problems such as generating zoning (Balling et al., 2004) or activity plans (Charypar and Nagel, 2005), transit network design (Bielli et al., 2002; Chakroborty, 2003), transit scheduling (Chakroborty et al., 1998), traffic parameter estimation (Sharma et al., 2004), dynamic traffic management (Lo et al., 2001; Abu-Lebdeh and Benekohal, 2003), and traffic incident detection (Srinivasan et al., 2000).

In signal timing design, GA has been used to optimize cycle time, green time, offsets, and stage sequences (e.g., Foy et al., 1992; Park et al., 2000; Park and Yun, 2005). These applications, however, are limited to problems in which no account is taken of re-routing in the optimization, and—with the possible exception of Park and Yun who consider a twelve-intersection network—are limited to small networks with few signalized intersections. In the network design context, Lee (1998), Taale and Van Zuylen (2003), and Ceylan and Bell (2004) used GA in optimizing signal timings, while anticipating re-routing impacts. For stage length and cycle time optimization (without considering offsets) to minimize total travel time, Lee (1998) presented a comparison of GA and simulated annealing with iterative and local search algorithms and showed that different algorithms perform better for different network supply and demand scenarios. Taale and Van Zuylen (2003) reported promising results from applying a GA to optimize green times within a NDP context, albeit on small artificial networks. In their approach to solve the NDP, Ceylan and Bell (2004) used the inverse of a system performance index, defined as the weighed sum of delays and stops for all traffic streams in the network, to optimize a common cycle time, green times, and offsets where route flows are constrained to a Stochastic User Equilibrium (Daganzo and Sheffi, 1977). Using a small network, they showed that a bi-level framework with GA gives more efficient results than an iterative algorithm in terms of system-wide travel costs. Although with a different application in mind, GA has also been proposed for addressing a number of problems with a similar bi-level structure, such as the analyses of Kim et al. (2001) and Stathopoulos and Tsekeris (2004) on the equilibrium-based OD matrix estimation problem; and those of Yin (2000) and Sumalee (2004) on bi-level NDPs.

Building on Ceylan and Bell's work we aim in the current article to present a GA-based signal timings optimization method that considers drivers rerouting and its application on a large-scale network. The optimum timings obtained are compared with those obtained from a method that does not consider rerouting. The impact of the choice of GA parameters on the performance of the algorithm is also presented. An illustrative objective function of total network-wide travel time will be adopted, though clearly many alternative objectives could be incorporated. A simulation-assignment model provides the junction delays based on which travel costs are calculated. Besides delays at signalized junctions, the model also enables the consideration of delays at nonsignalized junctions.

The next section presents the problem formulation and defines signal control design parameters. The methodology description is given in Section 3. The performance of the program is analyzed in Section 4, with conclusions given in the final section.

2 THE PROBLEM OF OPTIMIZING SIGNALS WITH EQUILIBRIUM CONSTRAINTS

The NDP of optimizing signal timings while anticipating drivers' rerouting is defined in the present section, and is hereafter referred to as "The Problem."

2.1 Problem formulation

The Problem is formulated as a mathematical program with equilibrium constraints (MPEC). The planner's upper level objective of minimizing total travel time[2] (TT) by altering signal timings is constrained so that the associated flows are at User Equilibrium based on the travel times resulting from the given signal timings. Additional signal setting feasibility constraints are also applied. The notation used is given in Table 1.

Table 1
Notation

L	Number of links
q_a	Flow on link a ($a = 1, 2, \ldots, L$)
T_{ij}	Total demand for travel between origin i and destination j
C	Common cycle time
θ_h	Offset at junction h, element of the vector of offsets θ
$\phi_{h,r}$	Duration of the green time for stage r at junction h, element of green time vector
$\psi(C, \theta, \phi)$	A vector of signal setting parameters (i.e., cycle time, offsets, and stage lengths)
$\mathbf{q}*(\psi)$	The vector of user equilibrium link flows given signal setting parameters ψ
f_{ijp}	Flow between origin i and destination j on route p
$t_a\,[\psi, \mathbf{q}*(\psi)]$	Travel time on link a
t_{ijp}	Total travel time on route p from origin i to destination j
TT	Total network travel time
S_h	Total number of stages at junction h
$I_{h,r}$	For junction h, the intergreen between the end of the green time for stage r and the start of the next green
G	Generation counter
ζ	Base 2 to base 10 converting function
ε	Total number of signal setting variables in a network
Ξ_m	The mth splice in a chromosome (e.g., Ξ_1 = the first 8 bits); see Section 3.1.
$\|P\|$	Population size
$\|G\|$	Generation number
c_w	Selection bias
Elite	Number of elite chromosomes

TT is defined as the sum of the product of the link flows and travel times over the whole network. At junctions each turning movement is represented by a separate link of zero-length, which has the associated delay put as travel time and included in the calculation of TT. TT is influenced by the signal timings, ψ, and link flow pattern $q^*(\psi)$ in the network (see Section 2.2).

Mathematically the problem is defined as:

$$\min_{\psi \in \Omega_o} \quad TT(\psi, q^*(\psi)) = \sum_{a=1}^{L} q_a t_a(\psi, q^*(\psi)) \tag{1}$$

$$\text{subject to:} \quad \psi(C, \theta, \phi) \in \Omega_o \tag{2}$$

$$\text{that is,} \quad C_{\min} \le C \le C_{\max} \tag{3}$$

$$0 \le \theta_h \le C - 1 \tag{4}$$

$$\phi_{h,r\,\min} \le \phi_{h,r} \le \phi_{h,r\,\max} \tag{5}$$

$$C = \sum_{r=1}^{S_h} \phi_{h,r} + \sum_{r=1}^{S_h} I_{h,r} \quad \forall\, h \tag{6}$$

and subject to the constraint that for a given Ψ, the UE flow vector $\mathbf{q}^*(\psi)$ is given by the variational inequality (Smith, 1979),

$$\mathbf{t}(\psi, \mathbf{q}^*) \cdot (\mathbf{q} - \mathbf{q}^*) \ge 0 \quad \forall \mathbf{q} \in \Upsilon \tag{7}$$

where $\mathbf{t}$ and $\mathbf{q}$ are the vectors of travel time function and link flow, respectively; and Υ is the feasible space of the link flow vector.

In this article, the traffic equilibrium problem is solved by use of the simulation-assignment modeling software package SATURN (Van Vliet, 1982). In contrast to conventional traffic assignment models, SATURN places great emphasis on the detailed representation of intersections, turning movements, and associated junction conflicts. Based on the principles of *cyclic flow profiles* (Robertson, 1969), its simulation sub-model determines junction delays at a number of flow states, based on junction turning demands estimated by the assignment sub-model. It uses this information to estimate marginal relationships between the travel time on each link/turn and the flow on that link/turn, and these functions are then passed back to the assignment sub-model for a re-computation of the route traffic demands, in the spirit of a "diagonalization" algorithm. The assignment sub-model uses the well-known Frank–Wolfe convex combinations method to determine the equilibrium flow pattern. Although for given travel time/flow functions the assignment sub-model has guaranteed convergence to an equilibrium solution, the process by which the assignment and simulation sub-models are alternately solved is a heuristic one. Nevertheless, the widespread use of this heuristic in practice has provided strong numerical evidence that good convergence can be assured through experimentation with the heuristic process.

2.2 Signal timing design variables and feasibility constraints

For a junction, Figure 1 shows the signal design variables: cycle time, offset, and the green time for the different stages. A phase is the set of movements that can take place simultaneously or the sequence of signal indicators received by such movements. The portion of the cycle over which a given combination of phases is given green is called a stage. For all signalized junctions, the offset—the difference in time of the start of the green between adjacent junctions—is defined from a common reference time for one (randomly chosen) stage. At a junction, the start of the green times for successive stages is defined by the inter-green period, measured from the end of the green of the preceding stage. For the problem formulated in Section 2.1, the technical feasibility of the design variables, which is ensured by Equations (3)–(6), is explained briefly in the following paragraphs.

To maintain signal coordination from cycle to cycle each junction in the area considered must operate with a *common cycle time* or a simple multiple of it (IHT, 1997). The common network cycle time, C, is constrained to a minimum, $C_{\min}$, and a maximum, $C_{\max}$, as shown in

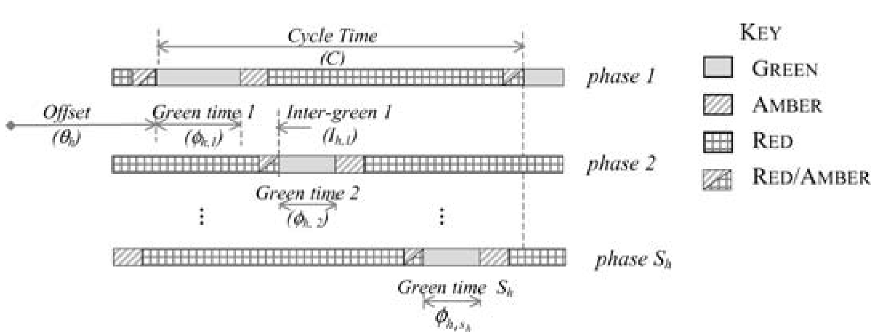


Fig. 1. Signal design variables for junction h.

Equation (3). $C_{\min}$ is determined by identifying that node which needs the longest duration just to accommodate the inter-green times and the minimum green times as shown in Equation (8). In the numerical tests reported later, $C_{\max}$ is constrained to 120 seconds.

$$C_{\min} = \operatorname{Max}\left\{\left(\sum_{r=1}^{S_h}\phi_{h,r\,\min} + \sum_{r=1}^{S_h} I_{h,r}\right): \quad h = 1, 2, \ldots, N\right\} \tag{8}$$

For any junction, h, the *offset* could only vary between zero and the cycle time minus one—see Equation (4).

For the numerical tests reported later in this study, a *green-time* duration of 7 seconds is used as a minimum for all stages in the network (Equation (9)). The maximum green time for a stage, $\phi_{h,r,\max}$, is obtained by assuming all the other stages at the junction just need the minimum green time. It is given by Equation (10), taking the lost time per cycle (calculated as the sum of the inter-green periods) and the minimum green times into account. Equation (10) ensures that the sum of the green times at a node together with the lost time in that cycle equals the total cycle time.

$$\phi_{h,r\,\min} = 7\,\text{sec} \quad \forall\, h, r \tag{9}$$

$$\phi_{h,r,\max} = C - \sum_{r=1}^{S_h} I_{h,r} - \sum_{\substack{y=1\\ y\neq r}}^{S_h} \phi_{h,y,\min} \tag{10}$$

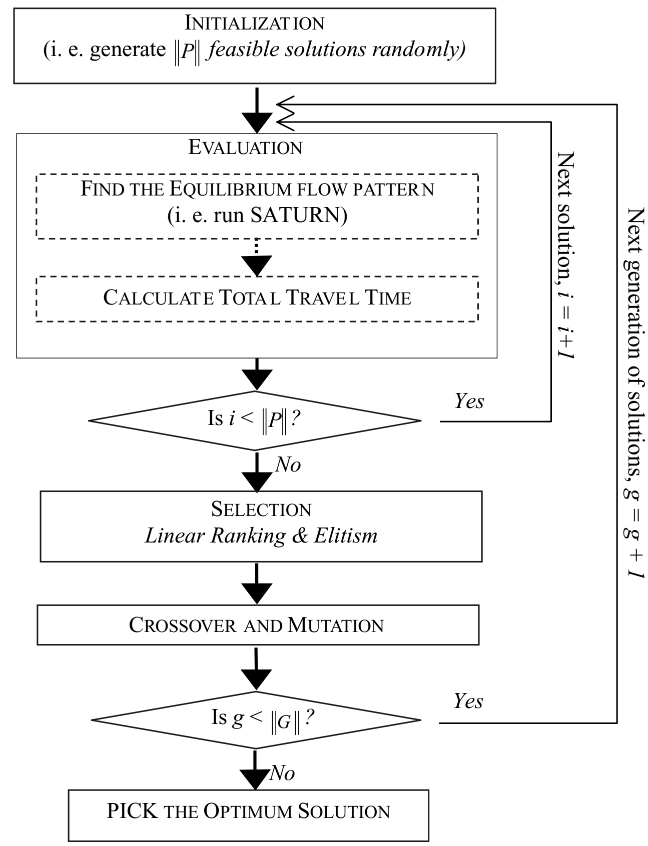


Fig. 2. GA-FITSUM flow chart.

3 SOLUTION METHODOLOGY AND IMPLEMENTATION

To solve the signal-setting NDP defined in Section 2, the problem will be cast in the form of a Genetic Algorithm (GA), whereby the GA aims to minimize the upper level objective TT with respect to the signal setting parameters, while maintaining flows at equilibrium for those signal settings (achieved through SATURN). Thus each time the objective function is evaluated at a particular choice of signal settings, a fully convergent run of SATURN is required to determine the flows that would arise under the given signal settings. Although the basic idea is straightforward, there are important issues concerned with the precise definition of the GA parameters and operators, and these are described in the sections below. As a precursor to these details, a brief description of the GA method and its terminology is provided.

GA is inspired by the theory of evolution. Initially, a population of chromosomes, each of which are potential solutions, are generated. GA evaluates each chromosome against an objective (fitness) function and, through a probabilistic selection process, selects some chromosomes to form what is known as an *intermediate population*. Mimicking the evolutionary strife for survival, the fitter chromosomes have higher probabilities of *selection*. Chromosomes from the intermediate population are then randomly paired to exchange genetic materials, and produce offspring in the *crossover* process. Lastly, in the *mutation* process, genes, on some probabilistically selected chromosomes, are made to mutate and form the next population. The process of going from one population to the next represents one generation in the execution of the GA. This evolutionary process goes on improving the fitness of the solutions through subsequent generations.

In this study, GA generates and improves on candidate signal timings for the network, for which the SATURN model determines the UE flow pattern and the respective total travel time, which is used for evaluating the sampled signal timings and improving the next batch of candidate timings. This is done repetitively until a "converged" solution is arrived at, or maximum number of generations is reached. GA-FITSUM (*Genetic Algorithm-based Formulation of Integrated Traffic Signal and User equilibrium Model*) is the name given to the computer program that solves this problem. A flow chart describing the implementation of the program is given in Figure 2.

Ξ_1	Ξ_2	Ξ_3	Ξ_4	Ξ_5	Ξ_6	Ξ_7
00101010	10101010	00101011	11010100	00010100	00100111	11111110
C	θ_1	θ_2	$\phi_{1,1}$	$\phi_{1,2}$	$\phi_{2,1}$	$\phi_{2,2}$

Fig. 3. Splices, chromosome, and represented variables—an illustrative example.

$$| C \,|\, \theta_1, \theta_2, \ldots, \theta_N \,|\, \phi_{1,1}, \phi_{1,2}, \ldots, \phi_{1,S_1} \,|\, \ldots \,|\, \phi_{N,1}, \ldots, \phi_{N,S_n} \,|$$

Fig. 4. Chromosome structure.

In the sections below, the details of the steps in the flow chart are described in greater depth.

3.1 Chromosome design

In the signal-setting NDP, the total number of decision variables, ε, is given by Equation (11). These are a common cycle time, N offsets (one for each junction), and S_m green times at each junction, giving a total of:

$$\varepsilon = 1 + N + \left(\sum_{m=1}^{N} S_m \right) \tag{11}$$

In the GA, each of these variables is represented by a splice of 8 bits—each bit taken from the binary set: $\{0, 1\}$. For example, considering a network with just two signalized junctions, each having two stages, the number of splices (ε) is 7. These splices will be combined to form a *chromosome* representing a vector of feasible signal setting parameters for the network. For the example mentioned above, the corresponding chromosome, splices, and represented variables are given in Figure 3. The general form of such a chromosome is shown in Figure 4 for a network with N signalized junctions.

3.2 Chromosome encoding and constraint representation

This section describes how the chromosome decoding is carried out, and how the constraints given by Equations (3)–(6) are handled in GA-FITSUM. The decoding scheme is based on Ceylan and Bell (2004).

a. Cycle time: is the proportion of the difference $C_{\max} - C_{\min}$ as defined by the first splice on the chromosome, plus $C_{\min}$.

$$C = C_{\min} + \frac{\zeta(\Xi_m)}{2^8 - 1}(C_{\max} - C_{\min}) \quad m = 1 \tag{12}$$

where: $\zeta(\Xi_m)$ = base 10 equivalent of splice m (Ξ_m);[3] $C_{\max} = 120$ seconds; and $C_{\min}$ is calculated using Equation (8).

b. Offset: for a stage at a junction h, it is the proportion of the cycle time as defined by splice m on the chromosome.

$$\theta_h = \frac{\zeta(\Xi_m)}{2^8 - 1}(C - 1) \quad m = 2, 3, \mathrm{K}, N + 1 \tag{13}$$

where: $h = m - 1$.

c. Green times: are defined as the sum of the minimum stage length and the proportion of the remaining green time $\phi_{h,r,\max} - \phi_{h,r,\min}$ as follows:

$$\phi_{h,r} = \phi_{h,r\min} + (\phi_{h,r\max} - \phi_{h,r\min}) \frac{\zeta(\Xi_m)}{\sum_{m=x+1}^{x+S_h} \zeta(\Xi_m)};$$

$$m = N + 2, N + 3, \ldots, \varepsilon \tag{14}$$

where: $x = N + 1 + \sum_{y=1}^{h-1} S_y$ is constant at junction h; $r = m - x$, is a stage at junction h; $h = 1, 2, \ldots, N$.

3.3 Initialization

The initial set of chromosomes (size of $\| P \|$) is randomly generated after network specific information such as the numbers of signalized junctions (N) and stages at each junction (S_h) have been extracted to determine ε [see Equation (11)].

3.4 Evaluation/fitness

Each chromosome is decoded and sent to the assignment model to obtain corresponding UE link flows and the associated total travel time, TT. GA-FITSUM then uses TT as the fitness function for the selection process.

3.5 Selection

In the *Selection* stage, GA-FITSUM uses the linear ranking approach proposed by Whitley (1989) for sampling the intermediate population, which will then be modified by GA operators (Sections 3.6 and 3.7). The chromosomes are first ranked in an ascending TT order and a "stochastic sampling with replacement" (Goldberg, 1989) is then carried out using a roulette wheel that uses probabilities, p_k, based on the rank, k, of each chromosome in a generation—see Equation (15). That is to say,

$$p_k = \frac{1}{\| P \|} \left(2 - c_w + (2c_w - 2) * \left(\frac{\| P \| - k}{\| P \| - 1} \right) \right) \tag{15}$$

where $c_w \in [1, 2]$ is the "selection bias"—higher values favor the better fit chromosomes during sampling.

As part of the selection procedure, *elitism* is applied which ensures the best performing chromosomes from the preceding generation are always included in the next generation without any alteration. The parameter *Elite* controls the number of such chromosomes that are passed to the next generation.

3.6 Crossover

During crossover, the elite chromosomes are made present in the breeding pool to share their good performing genes. Each chromosome in the intermediate population is assigned a probability of P_c to exchange its genetic materials. Those selected are randomly paired for which GA-FITSUM uses uniform crossover (Syswerda, 1989). Accordingly a mask chromosome with a random sequence of 0s and 1s is generated for each pair. The value of the mask chromosome determines which parent chromosome supplies the bit for a certain position in the chromosomes of their two offspring. The unselected chromosomes will be passed directly to the *mutation* process.

3.7 Mutation

The probability of mutation, P_m, determines the likelihood of mutation occurring on a certain chromosome. For those selected, the mutation is implemented by selecting a random point (bit) on the offspring's chromosome length and then changing the value of the that bit to 0 if it was 1, or vice versa.

4 CASE STUDY APPLICATION

GA-FITSUM's performance is considered here, both in terms of the GA search process and in terms of the optimality of the solution obtained in comparison with a traffic signal optimization process that does not anticipate rerouting. The evidence for this study arises from applications to the road network for the city of Chester in the UK, as illustrated in Figure 5. This network has 75 signalized junctions, 18 roundabouts, and 86 priority junctions. It caters for a total demand of 22,060 pcu/hour, generated from 132 zones.

4.1 GA parameters and performance

According to Whitley (1989), population diversity and selective bias are the two important issues influencing genetic search. As selective bias is increased—which could be implemented in GA-FITSUM by increasing

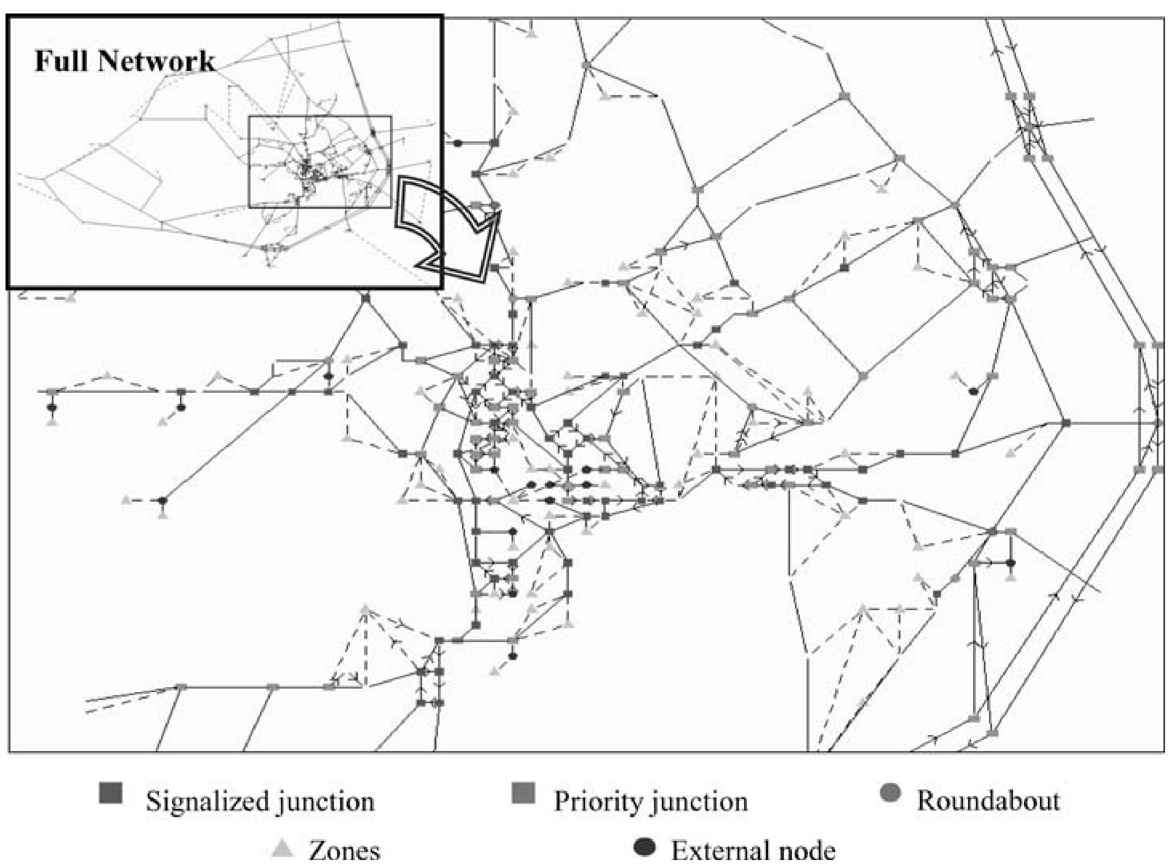


Fig. 5. Layout of the Chester network.

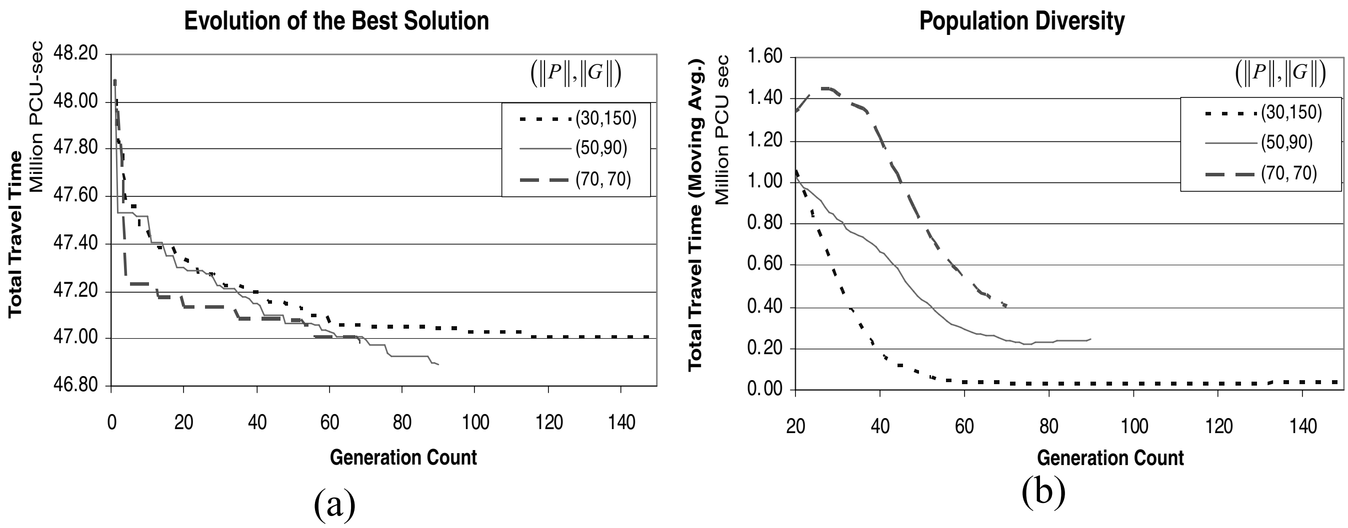


Fig. 6. Comparison of different ($\|P\|$, $\|G\|$) values. (a) Profile of improvement of best solution; (b) moving average (per 20 generations) of the variance in the total travel times of solutions in each generation.

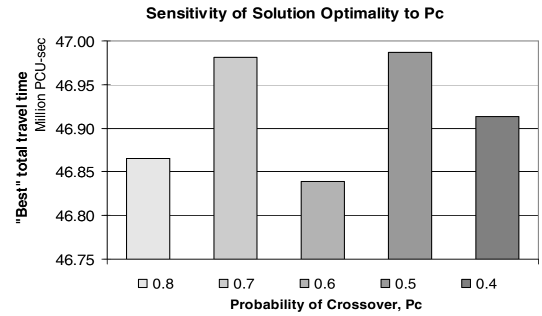
the *Elite* and c_w parameters in the selection process—the search focuses on the top individuals to exploit their best performing genes and leads to a fast convergence. A very high selective bias may result in a premature convergence. Low selective pressure, on the other hand, focuses on diversity and explorative search behavior. In GA-FITSUM such an effect could be implemented by higher values of $\|P\|$, P_m, and P_c. A higher $\|G\|$ would increase the number of chances GA gets to improve on the solution. Several researchers in the field of evolutionary optimization have tried to investigate the effect of the GA parameters so as to define these parameters optimally (see, e.g., Goldberg, 2002). Unfortunately, the most advanced result on optimal adjustment of GA parameters has been limited to very simple problems.

In this article the performance of GA is presented using the so-called "Evolution of Best Solution" (e.g., Figure 6a) and "Population Diversity" (e.g., Figure 6b) charts. The former shows the total travel time resulting from the best solution, in the vertical axis, against the generation that solution comes from on the horizontal axis. It indicates the speed of solution convergence. As the *Elitism* parameter is applied, the solution is observed to improve in successive generations. Charts of the latter kind show plots of the variance of the total travel time due to each solution in a generation, smoothed by averaging over a large number of (in this case 20) generations. It is chosen to indicate the possibility the search has to improve its solution by incorporating different genes.

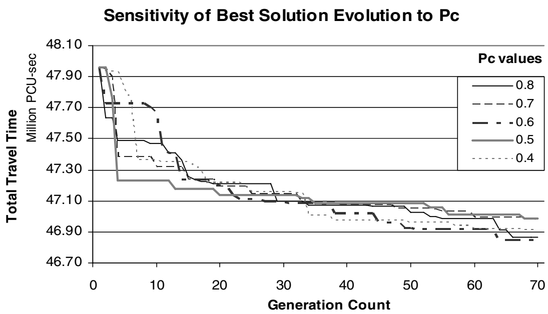
Figure 6 presents the comparison of different values of the population and generation number on the search process. The total number of chromosomes (solutions) evaluated has been constrained to about 4,500 so as to see whether it is better to have a larger population size and thus compromising the generation size, or vice versa. The other parameters were fixed as follows: $P_c = 0.5$, $P_m = 0.15$, $c_w = 1.2$, and *Elite* $= 1$. The population diversity is seen to decrease with lower values of $\|P\|$ in Figure 6b. The relationship between diversity and improvement in the best solution is clearly evident from the figure. For example, for $\|P\| = 30$ lower genetic diversity has limited the search process to fewer improvements after the 60th generation, whereas a high improvement rate is observed for $\|P\| = 70$ in the first few generations. High and sustained speed of convergence and a better solution are obtained with ($\|P\|$, $\|G\|$) equal to (50, 90).

The next parameter tested is the crossover parameter; see Figure 7. The values tested are: 0.8, 0.7, 0.6, 0.5, and 0.4. The values of $\|P\|$, $\|G\|$ and P_m are fixed at 70, 70 and 0.15, respectively. The test with $P_c = 0.6$ is shown to provide a better solution than the others. Similar patterns between solution improvement and diversity are also observed. The high population diversity up to the 30th generation is associated with faster best-solution improvement for $P_c = 0.5$. As shown in Figure 7c, the diversity decreases with GA evolution except for $P_c = 0.4$ for which diversity remains about constant after 40 generations.

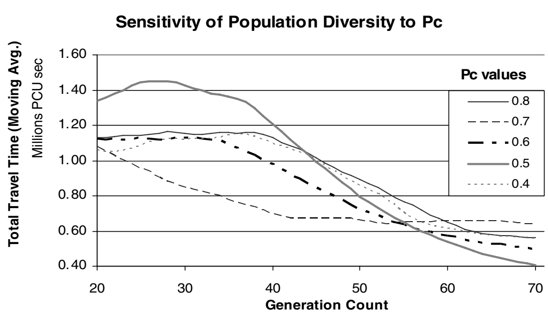
Similarly, the effect of the P_m parameter is presented in Figures 8a–c; values of $\|P\|$, $\|G\|$ and P_c are fixed at 70, 70, and 0.6, respectively. The best solution is found for $P_m = 0.25$. For the case of P_m, the effect of different P_m



(a) Optimality of final solution

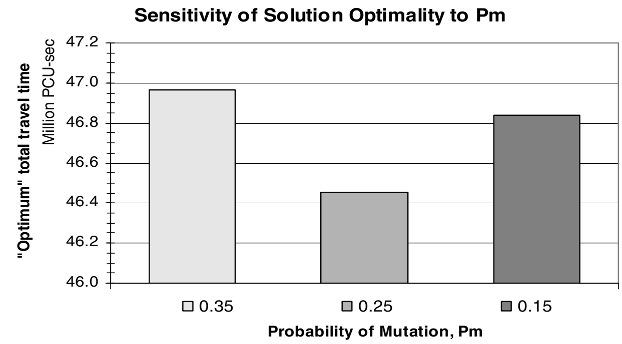


(b) Profile of improvement of best solution

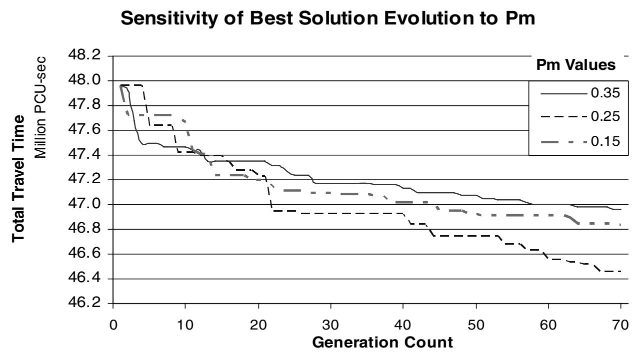


(c) Moving average (per 20 generations) of the variance in the fitness values of solutions in each generation

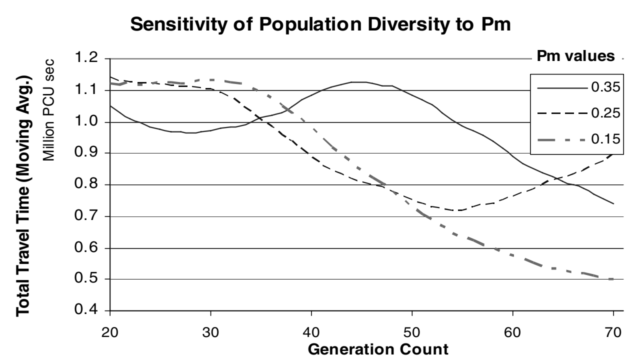
Fig. 7. Comparison of different P_C values with $\|P\| = 70$, $\|G\| = 70$, and $P_m = 0.15$.



(a) Optimality of final solution



(b) Profile of improvement of best solution



(c) Moving average (per 20 generations) of the variance in the fitness values of solutions in each generation

Fig. 8. Comparison of different P_m values with $\|P\| = 70$, $\|G\| = 70$, and $P_C = 0.6$.

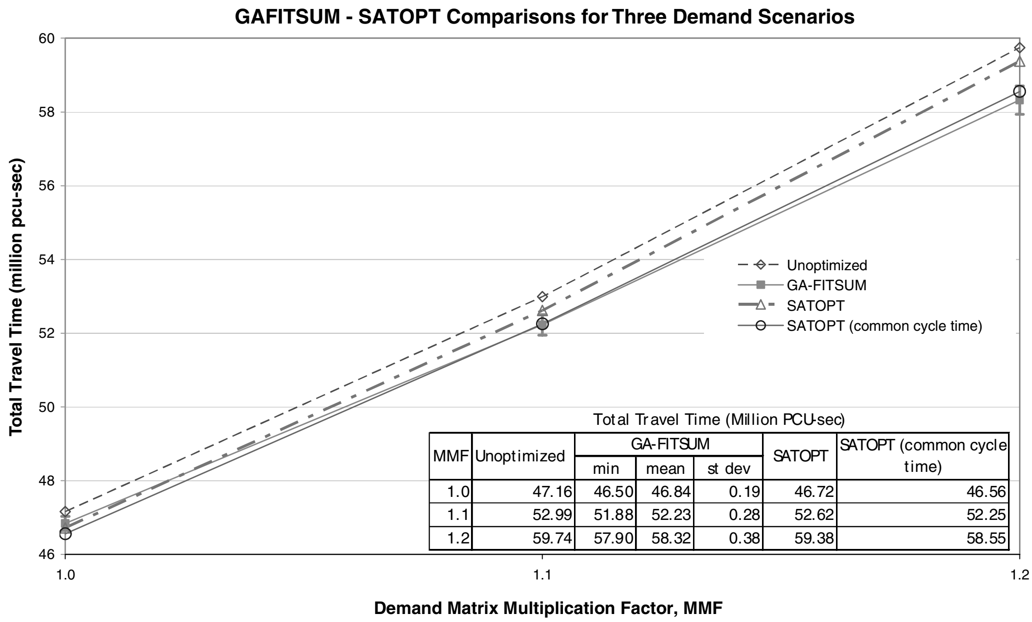
on the diversity of the solutions and the optimal solution found is, however, rather complex.

4.2 Local delay minimizing signal timings

For comparison, a SATURN-based signal optimization procedure, which will be referred to as SATOPT, was used. SATOPT optimizes stage lengths and offsets such that the total vehicle delay is minimized at each intersection using the SATURN simulation model. The method neither optimizes cycle time, nor considers the rerouting effect. After optimum SATOPT signal times are obtained, traffic is reassigned to take account of drivers' response to the new timings. The resulting total travel time is compared with that resulting from GA-FITSUM timings. In addition, to account for SATOPT not optimizing cycle times, and to check how stage timings and offset optimizations of GA-FITSUM compare with SATOPT, the best performing common cycle times from GA-FITSUM are also input and compared.

4.3 Results

The results from using GA-FITSUM for three demand scenarios are presented in Figure 9. The scenarios are defined by the matrix multiplication factors, MMF, applied on the current year OD matrix. To account for the effect of different initial random number generating



MMF	Unoptimized	GA-FITSUM min	GA-FITSUM mean	GA-FITSUM st dev	SATOPT	SATOPT (common cycle time)
1.0	47.16	46.50	46.84	0.19	46.72	46.56
1.1	52.99	51.88	52.23	0.28	52.62	52.25
1.2	59.74	57.90	58.32	0.38	59.38	58.55

Fig. 9. Total travel time comparison of SATOPT and GA-FITSUM.

Note: The error bars on mean TT from GA-FITSUM timings are for one standard deviation above and below the mean.

seeds, GA-FITSUM was run 10 times for each scenario. The mean TT and error bars representing one standard deviation above and below the mean are plotted in Figure 9. The inset table also shows the minimum TT obtained from GA-FITSUM. The TT from SATOPT optimized signal timings with and without the common cycle times is also presented. For benchmarking purposes, the TT from the un-optimized signal timings is also presented in Figure 9.

As shown in the figure, there does not seem to be much difference between GA-FITSUM and SATOPT for the first demand scenario. In fact, although the best solution from the 10 runs of the former gave the smallest TT of 46.5 million pcu-second, the mean TT from GA-FITSUM timings is bigger than that resulting from SATOPT timings. For higher levels of demand, however, it can be seen that GA-FITSUM gives better results than SATOPT.

SATOPT performs better when the common cycle time from the best performing GA-FITSUM timings are used, than when the initial node-specific signal timings are in operation. When compared with GA-FITSUM's best performing timings, SATOPT results in marginal increases in TT, which get bigger for higher levels of network congestion. This is because it does not consider the response of the traffic to the changed times.

To see the effect of not anticipating rerouting when optimizing signal times, the best performing signal times and, implicitly, link flows for the first demand scenario is given as input. For the optimum cycle time determined by GA-FITSUM, SATOPT calculated optimum stage lengths and offsets after which the traffic was reassigned. Figure 10 shows the difference in link flows as a result of the (new) SATOPT timings, which is unaccounted for in the conventional optimization calculations, which do not anticipate rerouting. For the case tested, the associated increase in TT is 0.5 million pcu-second per hour (which is around 25,000 pcu-hour of delay annually).

5 CONCLUSIONS

In this article, a GA-based method was presented for optimizing traffic signals in a way that anticipates rerouting of traffic, and its application on the city of Chester was presented. The GA-based method has given promising results in finding optimum signal timings with stable flows.

When compared to a local delay minimizing timings, the process by which routing is anticipated has given better results for the network considered in this article. The improvements after considering rerouting are relatively bigger when there is a higher level of congestion in the network. Future research should consider several traffic networks with varying levels of congestion, to explore the improvements associated with considering routing in optimizing signal timings.

A GA-based method should be run several times using different initial random seeds to better handle solution convergence issues. This will significantly increase the

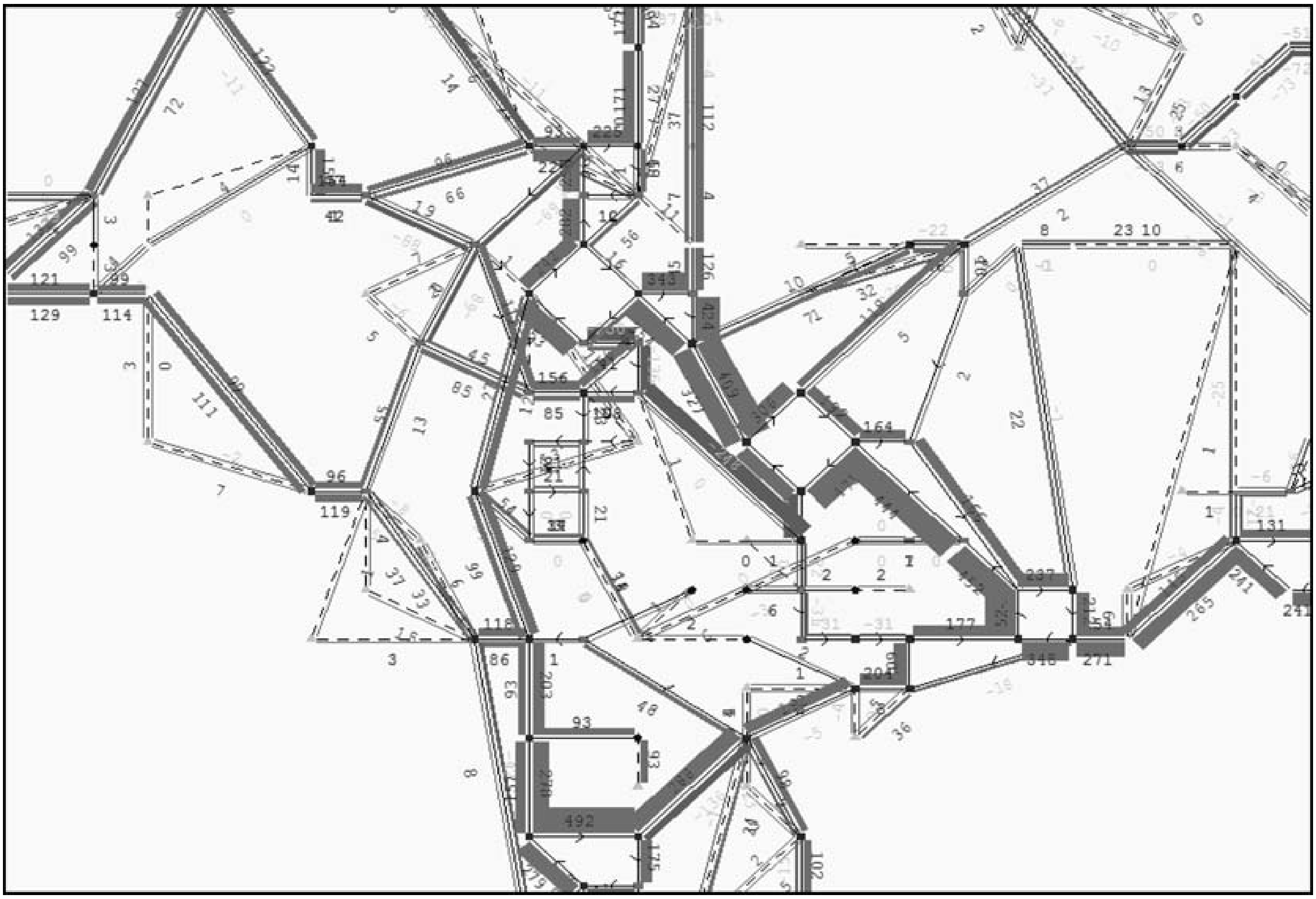


Fig. 10. Difference in link flows as a result of SATOPT timings (pcu/hour).

Note: The bands on the links show differences in hourly flows (positive or negative) due to SATOPT timings relative to those due to GA-FITSUM timings. The bandwidth is proportional to the magnitude of the difference.

relatively long computing time associated with the use of GA. The combination of population size, probabilities of crossover, and mutation as well as the chromosome design contributes to how GA performs. The question of how to find optimum values of the GA parameters is still a difficult issue and requires further research.

ACKNOWLEDGMENTS

This study was sponsored by the Norwegian Council of Universities' Committee for Development Research and Education, NUFU. We would like to acknowledge the detailed comments of all six anonymous referees, which helped us in improving an earlier version of this article.

NOTES

1. This may be contrasted with the case in which both the routing pattern and the signal settings are optimized with respect to a common system objective, which Varia and Dhingra (2004) solved with the use of a Genetic Algorithm. Although such a system of optimal allocation of the flows is a useful theoretical benchmark, it is not intended as a prediction of flows that could actually arise in practice.
2. It should be noted that the method described in this article could readily accommodate alternative objectives such as environmental pollution.
3. For a base 2 number, B, of 8 digits $(b_8\, b_7\, b_6 \ldots b_1)_2$ where $b_k \in \{0, 1\}, k = 1, 2, \ldots, 8$, the base 10 equivalent is given by, $\zeta\ B = \sum_{k=1}^{8} (b_k * 2^{k-1})$. It should be noted that $0 \leq \frac{\zeta B}{2^8 - 1} \leq 1$.

REFERENCES

Abdulaal, M. & LeBlanc, L. J. (1979), Continuous equilibrium network design models, *Transportation Research Part B*, **13**(1), 19–32.

Abu-Lebdeh, G. & Benekohal, R. F. (2003), Design and evaluation of dynamic traffic management strategies for congested conditions, *Transportation Research Part A*, **37**(2), 109–27.

Allsop, R. E. (1974), Some possibilities for using traffic control to influence trip distribution and route choice, in D. J. Buckley, ed., *Proceedings of the 6th International Symposium on Transport and Traffic Theory*, Elsevier, New York, 345–73.

Almond, J. & Lott, R. S. (1968), The Glasgow experiment: Implementation and assessment. *Road Research Laboratory Report* **142**, Road Research Laboratory, Crowthorne.

Balling, R., Powell, B. & Saito, M. (2004), Generating future land-use and transportation plans for high-growth cities using a genetic algorithm, *Computer-Aided Civil and Infrastructure Engineering*, **19**(3), 213–22.

Ben Ayed, O., Boyce, D. E. & Blair, C. E. III (1988), A general bi-level linear programming formulation of the network design problem, *Transportation Research Part B*, **22**(4), 311–8.

Bergendorff, P., Hearn, D. W. & Ramana, M. V. (1997), Congestion toll pricing of traffic networks, in P. M. Pardalos, D. W. Hearn, and W. W. Hager (eds.), *Lecture Notes in Economics and Mathematical Systems*, *450*: *Network Optimization*, Springer, Berlin, 51–71.

Bielli, M., Caramia, M. & Carotenuto, P. (2002), Genetic algorithms in bus network optimization, *Transportation Research Part C*, **10**(1), 19–34.

Ceylan, H. & Bell, M. G. H. (2004), Traffic signal timing optimisation based on genetic algorithm approach, including drivers' routing, *Transportation Research Part B*, **38**(4), 329–42.

Chakroborty, P. (2003), Genetic algorithms for optimal urban transit network design, *Computer-Aided Civil and Infrastructure Engineering*, **18**(3), 184–200.

Chakroborty, P., Deb, K. & Srinivas, B. (1998), Network-wide optimal scheduling of transit systems using genetic algorithms, *Computer-Aided Civil and Infrastructure Engineering*, **13**(5), 363–76.

Charypar, D. & Nagel, K. (2005), Generating complete all-day activity plans with genetic algorithms, *Transportation*, **32**(4), 369–97.

Chiou, S-W. (1999), Optimization of area traffic control for equilibrium network flows, *Transportation Science*, **33**(3), 279–89.

Cree, N. D., Maher, M. J. & Paechter, B. (1998), The continuous equilibrium optimal network design problem: A genetic approach, in M. G. H. Bell (ed.), *Transportation Networks: Recent Methodological Advances*, Pergamon, Oxford, 163–74.

Daganzo, C. F. & Sheffi, Y. (1977), On stochastic models of traffic assignment, *Transportation Science*, **11**(3), 253–74.

Dantzig, G. B., Harvey, R. P., Landsdown, Z. F., Robinson, D. W. & Maier, S. F. (1979), Formulating and solving the network design problem by decomposition, *Transportation Research Part B*, **13**(1), 5–17.

Foy, M., Benekohal, R. F. & Goldberg, D. E. (1992), Signal timing determination using genetic algorithms, *Transportation Research Record*, **1365**, 108–15.

Friesz, T. L., Cho, H. J., Mehta, N. J., Tobin, R. & Anandalingam, G. (1992), A simulated annealing approach to the network design problem with variational inequality constraints, *Transportation Science*, **26**, 18–26.

Friesz, T. L., Tobin, R. L., Cho, H. J. & Mehta, N. J. (1990), Sensitivity analysis based heuristic algorithms for mathematical programs with variational inequality constraints, *Mathematical Programming*, **48**, 265–84.

Goldberg, D. E. (1989), *Genetic Algorithms in Search, Optimisation, and Machine Learning*, Addison Wesley Longman Inc., Reading, MA.

Goldberg, D. E. (2002), *The Design of Innovation: Lesson from and for Competent Genetic Algorithms*, Kluwer Academic Publishers, Boston.

Hearn, D. W. & Ramana, M. V. (1998), Solving congestion toll pricing models, in S. Nguyen and P. Marcotte, (eds.), *Equilibrium and Advanced Transportation Modeling*, Kluwer Academic Publishers, Boston, pp. 109–24.

Hunt, P. B., Robertson, D. I., Bretherton, R. D. & Winton, R. I. (1981), SCOOT—A traffic responsive method of coordinating signals, *TRRL Laboratory Report* **1014**, TRRL, Berkshire, England.

(IHT) Institution of Highways & Transportation (1997), *Transport in the Urban Environment*. Institution of Highways & Transportation, London.

Kim, H., Baek, S. & Lim, Y. (2001), Origin-destination matrices estimated with a genetic algorithm from link traffic counts, *Transportation Research Record*, **1771**, 156–63.

Lawphongpanich, S. & Hearn, D. W. (2004), An MPEC approach to second-best toll pricing, *Mathematical Programming B*, **101**(1), 33–55.

LeBlanc, L. & Boyce, D. (1986), A bi-level programming for exact solution of the network design problem with user-optimal flows, *Transportation Research Part B*, **20**(3), 259–65.

Lee, C. (1998), *Combined traffic signal control and traffic assignment: Algorithms, implementation and numerical results*, Ph.D. dissertation, the University of Texas at Austin, Texas.

Lo, H. K., Chang, E. & Chan, Y. C. (2001), Dynamic network traffic control, *Transportation Research Part A*, **35**(8), 721–44.

Luo, Z., Pang, J. S. & Ralph, D. (1996), *Mathematical Programs with Equilibrium Constraints*, Cambridge University Press, New York.

Marcotte, P. (1981), An analysis of heuristics for the continuous network design problem, in V. F. Hurdle et al. (eds.), *Proceedings of the 8th International Symposium on Transportation and Traffic Theory*, University of Toronto Press, Toronto, 27–34.

Marcotte, P. (1986), Network design problem with congestion effects: A case of bi-level programming, *Mathematical Programming*, **34**, 142–62.

Meng, Q., Yang, H. & 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*, **35**(1), 83–105.

Park, B., Messer, C. J. & Urbanik, T. (2000), Enhanced genetic algorithm for signal timing optimization of oversaturated intersections, *Transportation Research Record*, **1727**, 32–41.

Park, B. & Yun, I. (2005). Stochastic optimization method for coordinated actuated signal systems. *Final Report of ITS Center Project: Evaluation of advanced traffic signal controllers*. Center for Transportation Studies, University of Virginia, USA.

Robertson, D. I. (1969), 'TRANSYT' method for area traffic control, *Traffic Engineering and Control*, **10**, 276–81.

Sharma, S., Lingras, P. & Zhong, M. (2004), Effect of missing values estimations on traffic parameters, *Transportation Planning and Technology*, **27**(2), 119–44.

Shepherd, S. P. & Sumalee, A. (2004), A genetic algorithms based approach to optimal toll level and location problems, *Networks and Spatial Economics*, **4**, 161–79.

Smith, M. J. (1979), The existence, uniqueness and stability of traffic equilibria, *Transportation Research Part B*, **13**(4), 295–304.

Srinivasan, D., Cheu, R. L., Poh, Y. P. & Ng, A. K. C. (2000), Development of an intelligent technique for traffic network incident detection, *Engineering Applications of Artificial Intelligence*, **13**(3), 311–22.

Stathopoulos, A. & Tsekeris, T. (2004), Hybrid meta-heuristic algorithm for the simultaneous optimization of the O-D trip matrix estimation, *Computer-Aided Civil and Infrastructure Engineering*, **9**(6), 421–35.

Steenbrink, P. A. (1974), *Optimization of Transport Networks*, John Wiley & Sons, London.

Sumalee, A. (2004), Optimal road user charging cordon design: A heuristic optimisation approach, *Computer-Aided Civil and Infrastructure Engineering*, **19**(5), 377–92.

Suwansirikul, C., Friesz, T. L. & Tobin, R. L. (1987), Equilibrium decomposed optimization: A heuristic for the continuous equilibrium network design problem, *Transportation Science*, **21**(4), 254–63.

Syswerda, G. (1989), Uniform crossover in genetic algorithms, in *Proceedings of the 3rd International Conference on Genetic Algorithms*, Morgan Kaufmann, 2–9.

Taale, H. & Van Zuylen, H. J. (2003). The Effects of Anticipatory Traffic Control for Several Small Networks. Paper presented at the 82nd Transportation Research Board Annual Meeting, Washington, DC, January 12–16, 2003.

Van Vliet, D. (1982), SATURN—A modern assignment model, *Traffic Engineering and Control*, **23**, 578–81.

Varia, H. R. & Dhingra, S. L. (2004), Dynamic optimal traffic assignment and signal time optimization using genetic algorithms, *Computer-Aided Civil and Infrastructure Engineering*, **19**, 260–73.

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), 707–29.

Wardrop, J. G. (1952), Some theoretical aspects of road traffic research, in *Proceedings of Institution of Civil Engineers*, **1**(2), 325–78.

Whitley, D. (1989), The GENITOR algorithm and selection pressure: Why rank based allocation of reproductive trial is best, in J. Schaffer, (eds.), *Proceedings of the 3rd International Conference on Genetic Algorithms*, Morgan Kaufmann, 116–21.

Webster, F. V. (1958), *Traffic Signal Settings*, Road Research Technical Paper No. 39, HMSO, London.

Yang, H. (1997), Sensitivity analysis for the elastic-demand network equilibrium problem with applications, *Transportation Research Part B*, **31**(1), 55–70.

Yin, Y. (2000), Genetic algorithms-based approach for bi-level programming models, *ASCE Journal of Transportation Engineering*, **126**(2), 115–20.