ELSEVIER

Transportation Research Part B 38 (2004) 329–342

TRANSPORTATION RESEARCH PART B

www.elsevier.com/locate/trb

Traffic signal timing optimisation based on genetic algorithm approach, including drivers' routing

Halim Ceylan [a,*], Michael G.H. Bell [b]

[a] *Department of Civil Engineering, Engineering Faculty, Pamukkale University, Denizli 20070, Turkey*

[b] *Department of Civil and Environmental Engineering, Imperial College, Exhibition Road, SW7 2BU London, UK*

Received 3 July 2002; received in revised form 6 December 2002; accepted 13 February 2003

Abstract

The genetic algorithm approach to solve traffic signal control and traffic assignment problem is used to tackle the optimisation of signal timings with stochastic user equilibrium link flows. Signal timing is defined by the common network cycle time, the green time for each signal stage, and the offsets between the junctions. The system performance index is defined as the sum of a weighted linear combination of delay and number of stops per unit time for all traffic streams, which is evaluated by the traffic model of TRANSYT [User guide to TRANSYT, version 8, TRRL Report LR888, Transport and Road Research Laboratory, Crowthorne, 1980]. Stochastic user equilibrium assignment is formulated as an equivalent minimisation problem and solved by way of the Path Flow Estimator (PFE). The objective function adopted is the network performance index (PI) and its use for the Genetic Algorithm (GA) is the inversion of the network PI, called the *fitness* function. By integrating the genetic algorithms, traffic assignment and traffic control, the GATRANSPFE (Genetic Algorithm, TRANSYT and the PFE), solves the equilibrium network design problem. The performance of the GATRANSPFE is illustrated and compared with mutually consistent (MC) solution using numerical example. The computation results show that the GA approach is efficient and much simpler than previous heuristic algorithm. Furthermore, results from the test road network have shown that the values of the performance index were significantly improved relative to the MC.
© 2003 Elsevier Ltd. All rights reserved.

* Corresponding author. Tel.: +90-258-2134030; fax: +90-258-2125548.
E-mail address: halimc@pamukkale.edu.tr (H. Ceylan).

0191-2615/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S0191-2615(03)00015-8

Nomenclature

$\boldsymbol{L}$	set of links on a road network, $\forall a \in \boldsymbol{L}$
$\boldsymbol{N}$	set of nodes, $\forall n \in \boldsymbol{N}$
$\boldsymbol{M}$	numbers of signal stages at a signalised road network
m	numbers of signal stages for a particular signalised junction, $\forall m \in \boldsymbol{M}$
c	a road network common cycle time
$c_{\min}$	minimum specified cycle time,
$c_{\max}$	maximum specified cycle time
θ	vector of feasible range of offset variables
ϕ	vector of duration of green times
I_i	intergreen time between signal stages
$\phi_{\min}$	minimum acceptable duration of the green indication for signal stage $\boldsymbol{\psi} = (c, \boldsymbol{\theta}, \phi)$ whole vector of feasible set of signal timings
$\boldsymbol{\Omega}_0$	vector of feasible region for signal timings
$\mathbf{q}$	vector of the average flow q_a on link a
$\mathbf{q}^*(\boldsymbol{\psi})$	vector of stochastic user equilibrium link flows
$\boldsymbol{W}$	set of origin–destination pairs
$\boldsymbol{P}_w$	set of paths each origin–destination pair w, $\forall w \in W$
$\mathbf{t}$	vector of origin–destination flows
$\mathbf{h}$	vector of all path flows
$\boldsymbol{\delta}$	link-path incidence matrix
$\boldsymbol{\Lambda}$	OD-path incidence matrix
$\mathbf{y}$	vector of expected minimum origin–destination cost
$\mathbf{g}(\mathbf{q}, \boldsymbol{\psi})$	vector of path travel times
$\mathbf{c}^0$	vector of free-flow link travel times
$\mathbf{c}(\mathbf{q}, \boldsymbol{\psi})$	vector of all link travel times
d_a^{U}	uniform delay at a signal-controlled junction
d_a^{ro}	random plus over saturation delay at a signalised junction
$\mathbf{K}$	matrix of link choice probabilities
k	the number of signal timing variables on a whole road network, the dimension of the problem is $k = \sum_{i=1}^{N} m_i + N$
$\mathbf{X}_{\mathrm{tt}}$	potential solution matrix of dimension $[\mathrm{NN} \times l]$ for the GA random search space
NN	population size
l	total number of binary bits in the string (i.e., chromosome)
tt	generation number

1. Introduction

In an urban road network controlled by fixed-time signals, there is an interaction between the signal timings and the routes chosen by individual road users. From the transportation engineering perspective, network flow patterns are commonly assumed fixed during a short period and control parameters are optimised in order to improve some performance index. On the other hand, from the transportation planning perspective, traffic assignment models are used to forecast network flow patterns, generally assuming that capacities decided by network supply parameters, such as signal settings, are fixed during a short period. The mutual interaction of these two processes can be explicitly considered, producing the so-called combined control and assignment problem.

The TRANSYT model proposed by Robertson (1969) has been widely recognised as one of the most useful tools in studying the optimisation of area traffic control. On the other hand, many traffic assignment models have been developed in order to find the link flows and path flows given origin–destination trip rates in an urban road network. One of these assignment models is the Path Flow Estimator (PFE). It has been developed by TORG (Transport Operation Research Group), in Newcastle University, to find link and path flows based on stochastic user equilibrium routing. However, it has been noted (Allsop, 1974; Gartner, 1974) that a full optimisation process needs to be applied where both problems are relevant; the area traffic control optimisation and user routing. The combined optimisation problem can be regarded as an Equilibrium Network Design Problem (ENDP) (Marcotte, 1983). Genetic Algorithms (GAs), first introduced by Goldberg (1989), have been applied to solve the ENDP (Lee and Machemehl, 1998; Cree et al., 1999; Yin, 2000).

A number of solution methods to this ENDP have been discussed and good results have been reported in a medium sized networks. Allsop and Charlesworth (1977) found mutually consistent traffic signal settings and traffic assignment for a medium size road network. In their study, the signal settings and link flows were calculated alternatively by solving the signal setting problem for assumed link flows and by carrying out the user equilibrium assignment for the resulting signal settings until convergence was achieved. The link performance function is estimated by evaluating delay for different values of flow and then fitting a polynomial function to these points. The resulting mutually consistent signal settings and equilibrium link flows, will, however, in general be non-optimal as has been discussed by Gershwin and Tan (1979) and Dickson (1981).

The first appearance of the GA for traffic signal optimisation was due to Foy et al. (1992), in which the green timings and common cycle time were the explicit decisional variables and the offset variables were the implicit decisional variable in a four-junction network when flows remain fixed. In the optimisation process, a simple microscopic simulation model was used to evaluate alternative solutions based on minimising delay. The results showed an improvement in the system performance when the GA was used and suggested that the GA has the potential to optimise signal timing. The results, however, were not compared with what could be achieved using existing optimisation tools. It was also concluded that the GA model may be able to solve more difficult problems than traditional control strategies and search methods in terms of convergence and that good convergence were reported in that study.

In this paper, for the purpose of solving the problem, a bi-level approach has been used. The upper level problem is signal setting while the lower level problem is finding equilibrium link flows based on the stochastic effects of drivers' routing. It is, however, known (see Sheffi and Powell, 1983) that there are local optima. It is not certain that the local solution obtained is also the global optimum because equilibrium signal setting is generally a non-convex optimisation problem. Hence, the GA approach is used to globally optimise signal setting at the upper level by calling TRANSYT (Vincent et al., 1980) traffic model to evaluate the objective function.

2. Formulation

The network performance index (PI) is a function of signal setting variables $\boldsymbol{\psi} = (c, \boldsymbol{\theta}, \boldsymbol{\phi})$ and equilibrium link flows $\mathbf{q}^*(\boldsymbol{\psi})$. The objective function is therefore to minimise PI with respect to equilibrium link flows $\mathbf{q}^*(\boldsymbol{\psi})$ subject to signal setting constraints. This gives the ENDP problem as the following minimisation problem:

$$\underset{\psi \in \boldsymbol{\Omega}_0}{\text{Minimise}} \quad \text{PI}(\psi, \mathbf{q}^*(\psi)) = \sum_{a \in \mathbf{L}} (Ww_a D_a(\psi, q^*(\psi)) + Kk_a S_a(\psi, q^*(\psi))) \tag{1}$$

$$\text{subject to} \quad \psi(c, \boldsymbol{\theta}, \boldsymbol{\phi}) \in \boldsymbol{\Omega}_0; \begin{cases} c_{\min} \leqslant c \leqslant c_{\max} & \text{cycle time constraints} \\ 0 \leqslant \theta \leqslant c & \text{values of offset constraints} \\ \phi_{\min} \leqslant \phi \leqslant \phi_{\max} & \text{green time constraints} \\ \sum_{i=1}^{m} (\phi_i + I_i) = c & \forall m \in \boldsymbol{M}, \forall n \in \boldsymbol{N} \end{cases}$$

where $\mathbf{q}^*(\boldsymbol{\psi})$ is implicitly defined by

$$\underset{q}{\text{Minimise}} \quad Z(\boldsymbol{\psi}, \mathbf{q}) \tag{2}$$

$$\text{subject to} \quad \mathbf{t} = \boldsymbol{\Lambda}\mathbf{h}, \quad \mathbf{q} = \boldsymbol{\delta}\mathbf{h}, \quad \mathbf{h} \geqslant \mathbf{0}$$

Then the fitness function (i.e., objective function for ENDP) becomes

$$\text{Maximise } F(\mathbf{x}) = \frac{1}{\text{PI}(\psi, \mathbf{q}^*(\psi))} \tag{3}$$

where $\text{PI}(\psi, \mathbf{q}^*(\boldsymbol{\psi}))$ is the value of the performance index of the network which is a function of equilibrium flow pattern $\mathbf{q}^*(\boldsymbol{\psi})$ and signal settings $\boldsymbol{\psi}$. All control variables are expressed in integer seconds, and $\mathbf{x}$ is a set of chromosomes that represents $\boldsymbol{\psi} \in \boldsymbol{\Omega}_0$, and F is a fitness function for the GA, to be maximised.

2.1. GA formulation for the upper-level problem

Suppose the fitness function (F) takes a set of $\boldsymbol{\psi}$ signal timing variables, $\boldsymbol{\psi} = (c, \boldsymbol{\theta}_1, \boldsymbol{\phi}_1, \ldots, \boldsymbol{\theta}_n, \boldsymbol{\phi}_n)$: $R^k \to R$. Suppose further that each decision variable $\boldsymbol{\psi}$ can take values from a domain $\boldsymbol{\Omega}_0 = [\boldsymbol{\psi}_{\min}, \boldsymbol{\psi}_{\max}] \subseteq R$ for all $\boldsymbol{\psi} \in \boldsymbol{\Omega}_0$. In order to optimise the objective function, we need to code the decision variables with some precision. The coding process is illustrated as follows:

$$\begin{array}{lllll} \text{Decision variables} & \psi = & |c| & |\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_n| & |\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_n| \\ \text{Mapping} & \downarrow & \downarrow & \downarrow & \downarrow \\ \text{Chromosome(string)} & \mathbf{x} = & |01010101| & |01010111, \ldots, 10101011| & |10101010, \ldots, 01010010| \end{array}$$

Then, the mapping from a binary string ($b_{l1} b_{l0} \ldots b_0$) representation of variables into a real numbers $\boldsymbol{\psi}$ from the range $[\boldsymbol{\psi}_{\min}, \boldsymbol{\psi}_{\max}]$ is carried out in following way:

(a) Convert the binary string $(b_{l1}b_{l0}\dots b_0)$ from base 2 to base 10:
(b)

$$(b_{l1}b_{l0}\dots b_0)_2 = \left(\sum_{j=l_i-1}^{l_i} b_j 2^j\right)_{10} = \Phi_i \quad i = 1,2,3,\dots,k \tag{4}$$

(c) Find a corresponding real number for each decision variable for a particular signal timing:

$$\psi_i = \psi_{i,\min} + \Phi_i \frac{\psi_{i,\max} - \psi_{i,\min}}{2^{l_i} - 1} \quad i = 1,2,3,\dots,k \tag{5}$$

where Φ_i is the integer resulting from (4). The decoding process from binary bit string to the real numbers is carried out by means of (5).

The following transformations are carried out for each signal timing variable for use in signal timing optimisation and traffic assignment purposes.

2.1.1. For common network cycle time

$$c = c_{\min} + \Phi_i \frac{(c_{\max} - c_{\min})}{2^{l_i} - 1} \quad i = 1 \tag{6}$$

2.1.2. For offsets

$$\theta_i = \Phi_i \frac{c}{2^{l_i} - 1} \quad i = 2,3,\dots,N \tag{7}$$

Mapping the vector of offset values to a corresponding signal stage change time at every junction is carried out as follows:

$$\theta_i = S_{i,j} \quad i = 1,2,\dots,N, \; j = 1,2,\dots,m$$

where $S_{i,j}$ is the signal stage change time at every junction.

2.1.3. For stage green timings

Let $p_1, p_2, \dots, p_i$ be the numbers representing by the genetic strings for m stages of a particular junction, and $I_1, I_2, \dots, I_m$ be the length of the intergreen times between the stages.

The binary bit strings (i.e., $p_1, p_2, \dots, p_i$) can be encoded as follows first;

$$p_i = p_{\min} + \Phi_i \frac{(p_{\max} - p_{\min})}{2^{l_i} - 1} \quad i = 1,2,\dots,m$$

where $p_{\min}$ and $p_{\max}$ are set as $c_{\min}$ and $c_{\max}$, respectively.

Then, using the following relation the green timings can be distributed to the all signal stages in a road network as follows second:

$$\phi_i = \phi_{\min,i} + \frac{p_i}{\sum_{k=1}^{m} p_i}\left(c - \sum_{k=1}^{m} I_k - \sum_{k=1}^{m} \phi_{\min,k}\right) \quad i = 1,2,\dots,m \tag{8}$$

2.2. The formulation for the lower-level problem

2.2.1. The (PFE) as a stochastic user equilibrium assignment (SUE)

The underlying theory of the PFE (see for details Bell et al., 1997) is the logit SUE model based on the notion that perceived cost determines driver route choice. The basic idea is to find the path flows and hence links flows, which satisfy an equilibrium condition where all travellers perceive the shortest path (allowing for delays due to congestion) according to their own perception of travel time.

The logit model assumes a particular distribution, the Gumbel distribution, for perceived travel times, which has the great advantage of allowing the formulation of a convex mathematical program whose solution is unique in the path flows. The equilibrium path flows are found by solving an equivalent minimisation problem. The attraction of the logit model is that it allows SUE flows and costs to be calculated by solving the convex mathematical program. The following minimisation problem

$$\text{minimise} \quad Z(\mathbf{h}) = \mathbf{h}^{\mathrm{T}}(\ln(\mathbf{h}) - \mathbf{1}) + \alpha \sum_{a \in L} \int_0^{q_a(\mathbf{h})} c_a(x)\,\mathrm{d}x \tag{9}$$

$$\text{subject to} \quad \mathbf{t} = \mathbf{\Lambda h}, \quad \mathbf{h} \geqslant \mathbf{0}$$

where all the notation is as previously stated, due to Fisk (1980), leads to a logit path choice model. Provided that the link cost functions are monotonically increasing with flows and assuming separable link cost functions, then $Z(\mathbf{h})$ is strictly convex and, since the constraints are also convex, it can be proved that there exists one unique solution to the program. The Kuhn–Karush–Tucker optimality conditions are

$$\nabla Z(\mathbf{h}) + \mathbf{\Lambda}^{\mathrm{T}}\mathbf{u}^* \geqslant \mathbf{0}, \quad \mathbf{h}^* \geqslant \mathbf{0} \quad \text{and} \quad (\nabla Z(\mathbf{h}) + \mathbf{\Lambda}^{\mathrm{T}}\mathbf{u}^*)^{\mathrm{T}}\mathbf{h}^* = \mathbf{0}$$

Since all paths are used

$$\nabla Z(\mathbf{h}) = \ln \mathbf{h} + \alpha \mathbf{g} \quad \text{As } \mathbf{h}^* > 0$$

$$\ln \mathbf{h}^* = -\alpha \mathbf{g}^* - \mathbf{\Lambda}^{\mathrm{T}}\mathbf{u}^*$$

It can be shown that this implies the following logit path choice model

$$h_p^* = t_w \frac{\exp(-\alpha g_p^*)}{\sum_{p \in P_w} \exp(-\alpha g_p^*)} \tag{10}$$

where t_w is the demand for origin destination pair w in $\boldsymbol{W}$, $\mathbf{u}$ is the dual variable and its definition can be obtained in Bell and Iida (1997). $\mathbf{u}$ may be associated with equilibrium link delays when it is divided by α, the dispersion parameter.

At the optimum

$$\nabla Z(\mathbf{h}^*) + \mathbf{\Lambda}^{\mathrm{T}}\mathbf{u}^* = \mathbf{0}$$

where $\nabla Z(\mathbf{h}^*) = \ln \mathbf{h}^* + \alpha \mathbf{g}^*$

At optimality

$$\ln \mathbf{h}^* + \alpha \mathbf{g}^* + \mathbf{\Lambda}^{\mathrm{T}}\mathbf{u}^* = \mathbf{0}$$

implying

$$h_p = \exp(-\alpha g_p - u_w).$$

From this logit model

$$\exp(-u_w) = \frac{t_w}{\sum_{p \in \boldsymbol{P}_w} \exp(-\alpha g_p)}$$

so

$$-u_w = \ln t_w - \ln \sum_{p \in \boldsymbol{P}_w} \exp(-\alpha g_p)$$

Hence

$$-\mathbf{u} = \ln \mathbf{t} + \alpha \mathbf{y}$$

PFE consists of two loops; an outer loop, which generates paths, and an inner loop, which assigns flows to paths according to logit path choice model (10). The SUE path flows are found by solving an equivalent optimisation problem (9) iteratively. An outer loop generates paths and an inner loop assigns flows to paths according to a logit path choice model.

2.3. The GATRANSPFE solution of the ENDP

A decoded genetic string is required to translate into the form of TRANSYT and PFE inputs, where TRANSYT model accepts the green times as stage start times, hence offsets between signal-controlled junctions, and the PFE requires the cycle time and duration of stage greens for that stages. The assignment of the decoded genetic strings to the signal timings is carried out using the following relations in the GATRANSPFE.

1. *For road network common cycle time*

$$c \leftarrow p(i,j) \quad i = 1, \;\; j = 1, 2, 3, \ldots, \mathrm{NN}$$

where p represents the corresponding decoded parent chromosome, j represents the population index, and i represents the first individual in the chromosome set.

2. *For offset variables*

$$\theta_n(i,j) \leftarrow p(i,j), \quad i = 2, 3, 4, \ldots, N, \;\; j = 1, 2, 3, \ldots, \mathrm{NN},$$

Since there is no closed-form mapping for offset variables, it is common to map these values to the interval $(0, c)$, hence offset values are mapped using (7). The decoded offset values are in some cases higher than the network cycle time due to the coding process in the GA. In this case, the remainder of a division between $p(i,j)$ and the c (i.e., modulo division) is assigned as a stage change time as follows:

$$\theta_n(i,j) \leftarrow \mathrm{MOD}(p(i,j), c), \;\; i = 2, 3, 4, \ldots, N \;\; j = 1, 2, 3, \ldots, \mathrm{NN}$$

3. *For green timing distribution to signal stages as a stage change time is*

$$\theta_{n,m}(i,j) = \theta_{n,m-1}(i,j) + ((I + \phi)_{n,m}(i,j)) \leqslant c; \quad \forall n \in \boldsymbol{N}, \;\; \forall m \in \boldsymbol{M}, \;\; i = 1, 2, 3, \ldots, M$$

The solution steps for the GATRANSPFE is:

Step 0. Initialisation. Set the user-specified GA parameters; represent the decision variables ψ as binary strings to form a chromosome **x** by giving the minimum $\psi_{\min}$ and maximum $\psi_{\max}$ specified lengths for decision variables.
Step 1. Generate the initial random population of signal timings $\mathbf{X}_{\text{tt}}$; set tt = 1.
Step 2. Decode all signal timing parameters of $\mathbf{X}_{\text{tt}}$ by using (6)–(8) to map the chromosomes to the corresponding real numbers.
Step 3. Solve the lower level problem by way of the PFE. This gives a SUE link flows for each link a in $\boldsymbol{L}$.

At Step 3, the link travel time function adapted for the PFE is the sum of free-flow travel time under prevailing traffic conditions (i.e., c_a^0) and average delay to a vehicle at the stop-line at a signal-controlled junction by simplifying the offset expressions for the PFE link travel time function, where the appropriate expressions for the delay components can be obtained in Ceylan (2002), as follows:

$$c_a(q_a, \psi) = c_a^0 + d_a^{\text{U}} + d_a^{\text{ro}}$$

Step 4. Get the network performance index for resulting signal timing at Step 1 and the corresponding equilibrium link flows resulting in Step 3 by running TRANSYT.
Step 5. Calculate the fitness functions for each chromosome x_j using the expression (3).
Step 6. Reproduce the population $\mathbf{X}_{\text{tt}}$ according to the distribution of the fitness function values.
Step 7. Carry out the crossover operator by a random choice with probability P_c.
Step 8. Carry out the mutation operator by a random choice with probability P_m, then we have a new population $\mathbf{X}_{\text{tt}+1}$.
Step 9. If the difference between the population average fitness and population best fitness index is less than 5%, re-start population and go to the Step 1. Else go to Step 10.
Step 10. If tt = maximal generation number, the chromosome with the highest fitness is adopted as the optimal solution of the problem. Else set tt = tt + 1 and return to Step 2.

3. Numerical application

A test network is chosen based upon the one used by Allsop and Charlesworth (1977) and Chiou (1998). Basic layouts of the network and stage configurations for GATRANSPFE are given in Fig. 1a and b, where Fig. 1a is adapted from Chiou (1998) and Fig. 1b adopted from Charlesworth (1977). Travel demands for each origin and destination are those used by Charlesworth (1977) and also given in Table 1. This numerical test includes 20 origin–destination pairs, and 21 signal setting variables at six signal-controlled junction.

The GATRANSPFE is performed with the following user-specified parameters:

Population size is 40.
Reproduction operator is binary tournament selection.

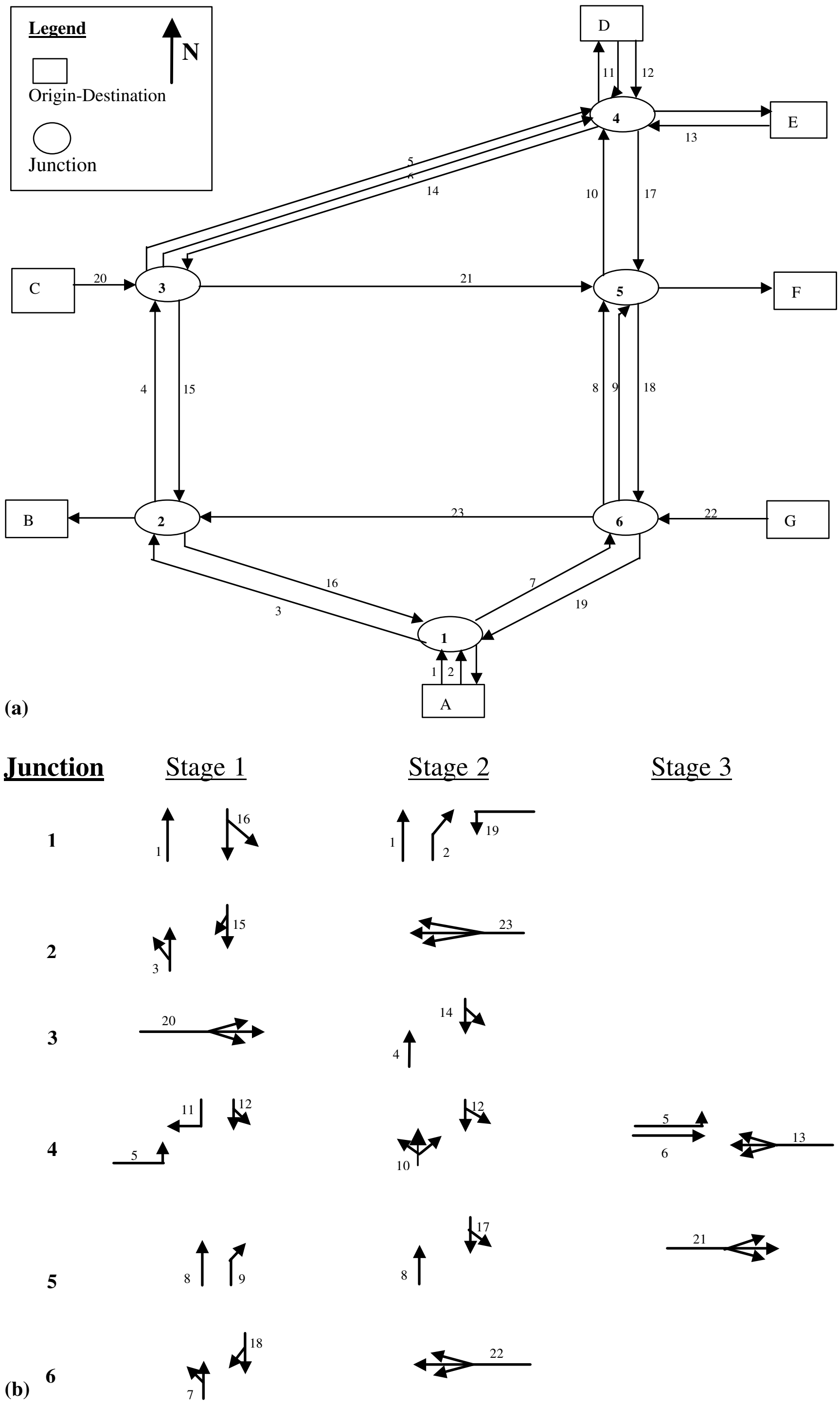

Fig. 1. (a) Layout for Allsop and Charlesworth's test network. (b) Stage configurations for the test network.

Table 1
Travel demand for Allsop and Charlesworth's network in vehicles/h

Origin/destination	A	B	D	E	F	Origin totals
A	–	250	700	30	200	1180
C	40	20	200	130	900	1290
D	400	250	–	50[a]	100	800
E	300	130	30[a]	–	20	480
G	550	450	170	60	20	1250
Destination totals	1290	1100	1100	270	1240	5000

[a] Where the travel demand between O–D pair D and E are not included in this numerical test which can be allocated directly via links 12 and 13.

Crossover operator is uniform crossover, and the probability is 0.5.
Mutation operator is creep mutation operator, and the probability is 0.02.
The maximal number of generation is 100.

The signal timing constraints are given as follows:

$c_{\min}, c_{\max} = 36, 120$ s	Common network cycle time
$\theta_{\min}, \theta_{\max} = 0, 120$ s	Offset values
$\phi_{\min} = 7$ s	Minimum green time for signal stages
$I_{1-2}, I_{2-1} = 5$ s	Intergreen time between the stages

3.1. GATRANSPFE solution for Allsop and Charlesworth's network

Although the bi-level problem (1) is non-convex and only a local optimum is expected to be obtained, in this numerical test, the GATRANSPFE model is able to avoid being trapped in a bad local optimum. The reason for this is that the model starts with a large base of solutions, each of which is pushed to converge to the optimum. If there is no more improvement on the population best fitness and population average fitness for the current generation, the GATRANSPFE re-starts the population. This has the effect of jumping from the current hill to different hills. The method applied is not dependent on the initial assignments of the signal setting variables. The random number seed given controls the initial sets of solutions within the population size. Unlike the GATRANSPFE solution of the network, the MC solution requires the initial assignment.

The application of the GATRANSPFE model to Allsop and Charlesworth's network can be seen in Fig. 2, where the convergence of the algorithm and improvement on the network performance index and hence the signal timings can be seen. The model calculates the fitness of each individual chromosome x_j in the population. The maximum fitness value found in the current generation is noted, then for each population pool, the selection, crossover and mutation operators are applied. When the differences between the population average fitness and population best fitness of the current generation is less than 5% then the algorithm re-starts with new randomly generated parents, whilst keeping the best fit chromosome from the previous population. The reason for this is to improve the speed of the model towards the optimum.

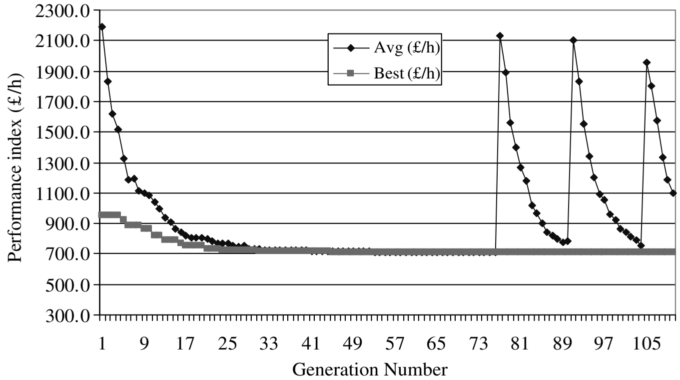

Fig. 2. The application of GATRANSPFE model to the test network.

In Fig. 2, there are no improvements on the best fitness value on the first few generations. The reason for this is that in the first iterations, the algorithm finds a chromosome with very good fitness value which is better than average fitness of the population. The algorithm keeps the best fitness then starts to improve population average fitness to the best chromosome while improving the best chromosome to optimum or near optimum. The considerable improvement on the objective function usually takes place in the first few iteration because the GA start with randomly generated chromosomes in a large population pool. After that, small improvements to the objective function takes place since the average fitness of the whole population will push forward the population best fitness by way of genetic operators, such as mutation and crossover.

Model analysis is carried out for the 75th generation, where the difference between the population average fitness and population best fitness is less than 5%, and network performance index obtained for that generation is 712.5 £/h. The model convergence can be seen in Fig. 3. The restart process began after the 75th generation and there was not much improvement to the population best fitness previously found as can be seen in Fig. 2.

Table 2 shows the signal timings and the final value of the performance index in terms of £/h and veh-h/h. The common network cycle time resulting from the GATRANSPFE application is 77 s and the start of greens for every stage in the signalised junctions are presented in Table 2.

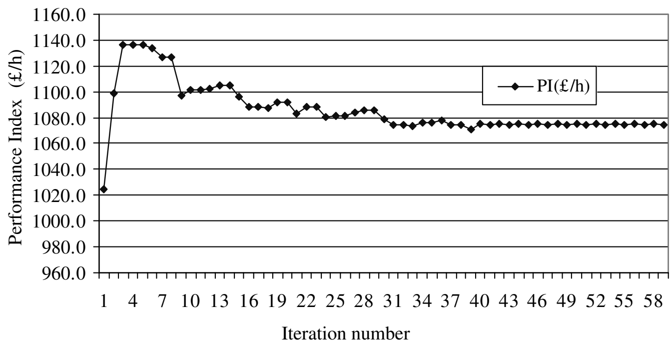

Fig. 3. The convergence behaviour of MC calculation.

Table 2
The final values of signal timings derived from the GATRANSPFE model

Performance index		Cycle time c (s)	Junction number i	Start of green in seconds		
£/h	veh-h/h			Stage 1 $S_{i,1}$	Stage 2 $S_{i,2}$	Stage 3 $S_{i,3}$
712.5	75.4	77	1	0	32	–
			2	59	25	–
			3	13	60	–
			4	44	72	20
			5	64	5	30
			6	47	6	–

3.2. MC solution for Allsop and Charlesworth's network

The MC calculations were carried out with the initial set of signal timings given in Table 3, where the signal timing are equally distributed to the signal stages. For this initial set of signal timings, along with equilibrium link flows resulting from the PFE, the initial performance index and its corresponding value of veh-h/h is given in Table 3.

As can be seen in Fig. 3, for the first iteration after performing a full TRANSYT run with the corresponding equilibrium flows, the value of the performance index increased from 1024 £/h to 1100 £/h, an increase of 7%. In the second iteration the MC soluton also increases the system performance index. Thereafter, alternately carrying out the two separate procedures of traffic assignment and TRANSYT optimisation of the signal timings, the final value of the performance index is 1075 £/h. This shows that the MC solution increases the system performance index by 5% when it is compared to the initial value of 1024 £/h. Fluctuations of the value of the performance index from iteration to iteration is obvious, which shows the non-optimal characteristic of the mutually consistent signal settings and equilibrium flows for the solution of the bi-level problem. The total number of iterations in performing the MC calculations in Fig. 3 is 60. The maximum degree of saturation is 0.97.

Table 4 shows the final values of the start of green timings for each signalised junction and performance index resulting from the MC solution. The network common cycle time is 82 s.

As for the solution of the MC calculation, Fig. 3 showed that the MC calculation increases the system performance index. In terms of the convergence, the MC is dependent on the ini-

Table 3
Initial signal timing assignment for use in the MC

Performance index		Cycle time c (s)	Junction number i	Start of green in seconds		
£/h	veh-h/h			Stage 1 $S_{i,1}$	Stage 2 $S_{i,2}$	Stage 3 $S_{i,3}$
1024.0	110.0	70	1	0	35	–
			2	0	35	–
			3	0	35	–
			4	0	23	46
			5	0	23	46
			6	0	35	–

Table 4
The final values of signal timings resulting from the MC solution

Performance index		Cycle time c (s)	Junction number i	Start of green in seconds		
£/h	veh-h/h			Stage 1 $S_{i,1}$	Stage 2 $S_{i,2}$	Stage 3 $S_{i,3}$
1075.0	116.0	82	1	32	72	–
			2	15	66	–
			3	52	20	–
			4	2	32	59
			5	27	62	5
			6	80	46	–

tial assignment. Various sets of initial signal timings were used as a starting point for the MC (see Ceylan, 2002). Only one set of the initial solutions converged to the predetermined threshold value that is presented in Fig. 3.

4. Conclusions

1. Allsop and Charlesworth's example network was used as an illustrative example for showing the performance of the GATRANSPFE method in terms of resulting values of performance index for the whole network and the degree of saturation on links. The performance of the proposed method in solving the non-convex bi-level problem showed in that the differences between resulting values of performance index in all cases were negligible at the 75th generation. Furthermore, none of the degree of saturation, resulting from the GATRANSPFE model, were over 90%. The GATRANSPFE model showed good improvement over the MC calculation in terms of the final values of performance index, with a 34% improvement over the MC solution of the problem, at the 75th generation, and an improvement in terms of convergence in all cases.
2. The MC solution of the problem was dependent on the initial set of signal timings and its solution was sensitive to the initial assignment. Depending on the initial signal timings, the convergence of the MC solution was not guaranteed. Note that applying the MC solution to Allsop and Charlesworth's road network caused the network performance index to increase compared with the initial performance index, whilst the GATRANSPFE model converged to the optimal solution (i.e., at the 75th generation) irrespective of the initial signal timings.
3. As for the computation efforts for the GATRANSPFE model, performed on PC 166 Toshiba machine, each iteration for this numerical example was less than 16.5 s of CPU time in Fortran 90. The total computation efforts for complete run of the GATRANPFE model run was 18.4 h. On the other hand, the computation effort for the MC solution on the same machine was performed for each iteration in less than 20 s of CPU time and the complete run did not exceed 1 h on that machine.
4. In this work, the effect of the stage ordering to a network performance index is not taken into account due to the coding procedure of the GATRANSPFE. Future work should take into account the effect of the stage orders by appropriately representing the stage sequences as a suitable GA code.

Acknowledgements

The author would like to thank to David Carroll who has provided me some part of the source code of the genetic algorithm. The work reported here was sponsored by the scholarship of Pamukkale University, Turkey. We thank the anonymous referees for their constructive and useful comments on the paper.

References

Allsop, R.E., 1974. Some possibilities of using traffic control to influence trip distribution and route choice. In: Transportation and Traffic Theory (Proceedings of the Sixth International Symposium on Transportation and Traffic Theory). Elsevier, Amsterdam, pp. 345–374.

Allsop, R.E., Charlesworth, J.A., 1977. Traffic in a signal-controlled road network: an example of different signal timings including different routings. Traffic Engineering Control 18 (5), 262–264.

Bell, M.G.H., Iida, Y., 1997. Transportation Network Analysis. John Wiley and Sons, Chichester, UK.

Bell, M.G.H., Shield, C.M., Busch, F., Kruse, G., 1997. Stochastic user equilibrium path flow estimator. Transportation Research 5C, 197–210.

Ceylan, H., 2002. A genetic algorithm approach to the equilibrium network design problem. Ph.D. Thesis, University of Newcastle upon Tyne.

Charlesworth, J.A., 1977. The calculation of mutually consistent signal settings and traffic assignment for a signal-controlled road network. In: Proceedings of the Seventh International Symposium on Transportation and Traffic Theory. Elsevier, Kyoto, pp. 545–569.

Chiou, S.-W., 1998. Optimisation of area traffic control for equilibrium network flows. Ph.D. Thesis, University College, London.

Cree, N.D., Maher, M.J., Paechter, B., 1999. The continuous equilibrium optimal network design problem: a genetic approach. Proceedings of the Fourth Euro Transportation Meeting. pp. 163–174.

Dickson, T.J., 1981. A note on traffic assignment and signal timings in a signal-controlled road network. Transportation Research 15B (4), 267–271.

Fisk, C., 1980. Some developments in equilibrium traffic assignment. Transportation Research 14B, 243–255.

Foy, M.D., Benekohal, R.F., Goldberg, D.E., 1992. Signal timing determination using genetic algorithms. In: Transportation Research Record 1365 TRB. National Research Council, Washington, DC, pp. 108–115.

Gartner, N.H., 1974. Area traffic control and network equilibrium methods. In: Floran, M. (Ed.), Traffic Equilibrium Methods. Springer-Verlag, Berlin, pp. 274–297.

Gershwin, S.B., Tan, H.N., 1979. Hybrid optimisation: optimal static traffic control constrained by drivers' route choice behaviour. Laboratory for Information and Decision System Report LIDS-p-870, Massachusetts Institute of Technology.

Goldberg, D.E., 1989. Genetic Algorithms in Search, Optimisation and Machine Learning. Addison-Wesley, Harlow, England.

Lee, C., Machemehl, R.B., 1998. Genetic algorithm, local and iterative searches for combining traffic assignment and signal control. Traffic and transportation studies. In: Proceedings of ICTTS 98. pp. 489–497.

Marcotte, P., 1983. Network optimisation with continuous control parameters. Transportation Science 17 (2), 181–197.

Robertson, D.I., 1969. TRANSYT: a traffic network study tool. RRL Report, LR 253, Transport and Road Research Laboratory, Crowthorne.

Sheffi, Y., Powell, W., 1983. Optimal signal settings over transportation networks. Journal of Transportation Engineering 109 (6), 824–839.

Vincent, R.A., Mitchell, A.I., Robertson, D.I., 1980. User guide to TRANSYT, version 8. TRRL Report LR888, Transport and Road Research Laboratory, Crowthorne.

Yin, Y., 2000. Genetic-algorithm-based approach for bilevel programming models. Journal of Transportation Engineering 126 (2), 115–120.