ELSEVIER

Transportation Research Part B 39 (2005) 169–185

TRANSPORTATION
RESEARCH
PART B

www.elsevier.com/locate/trb

Genetic algorithm solution for the stochastic equilibrium transportation networks under congestion

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

[a] *Department of Civil Engineering, Engineering Faculty, Pamukkale University, Denizli 20017, Turkey*
[b] *Department of Civil and Environmental Engineering, Imperial College, Exhibition Road, SW7 2BU London, UK*

Received 6 September 2002; received in revised form 2 December 2003; accepted 14 April 2004

Abstract

A bi-level and mutually consistent (MC) programming techniques have previously been proposed, in which an area traffic control problem (ATC) is dealt with as upper-level problem whilst the users' equilibrium traffic assignment is dealt with as lower-level problem. In this study, genetic algorithm (GA) approach has been proposed to solve upper-level problem for a signalized road network under congestion. Stochastic user equilibrium (SUE) traffic assignment is applied at the lower-level. At the upper-level, GA provides a feasible set of signal timings within specified lower and upper bounds on signal timing variables and feeds into lower-level problem. The SUE assignment is solved by way of Path Flow Estimator (PFE) and TRANSYT traffic model is applied at upper-level to obtain network performance index (PI) and hence fitness index. Network performance index is defined as the sum of a weighted linear combination of delay and number of stops per unit time under various levels of traffic loads. For this purpose, the genetic optimizer, referred to as GATRANSPFE, combines the TRANSYT model, used to estimate performance, with the PFE logit assignment tool, used to predict traffic reassignment, is developed. The GATRANSPFE that can solve the ATC and SUE traffic assignment problem has been applied to the signalized road networks under congestion. The effectiveness of the GATRANSPFE over the MC method has been investigated in terms of good values of network performance index and convergence. Comparisons of the performance index resulting from the GATRANSPFE and that of mutually consistent TRANSYT-optimal signal settings and SUE traffic flows are made.

Keywords: Genetic algorithm; Bi-level programming; Mutually consistent; Stochastic user equilibrium

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

0191-2615/$ - see front matter
doi:10.1016/j.trb.2004.04.001

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. The problem falls within the framework of a leader–follower or Stackelberg game, where the supplier is the leader and the user is the follower (Fisk, 1984). When drivers follow Wardrop's (1952) first principle, (i.e. user equilibrium), the problem is called the "equilibrium network design problem" (ENDP), which is normally non-convex. In the case of an over demand, the road users will use less costly routes in which drivers will perceive minimum path travel time. The application of SUE traffic assignment is to be more attractive than the user equilibrium (UE) for which the road users have perfect information on the road system for the ENDP. Under UE condition only less costly routes will carry flow but under SUE only less costly routes which are perceived as a cheaper routes by road users will carry flow.

The congestion plan is an important role in an urban transportation networks especially in peak times, where demand is excessive from the capacity. The signal timings and the stochastic user equilibrium (SUE) link flow for the ENDP comes a critical stage when the signalized road networks are overloaded. Critical stage at which demand excessive from capacity needs to be considered by means of re-optimizing the signal timings with SUE link flows. An increase in demand will cause in extra green timings at a signalized junction, and requires possibly better co-ordinations between the signalized junctions.

In dealing with the dependence of the stochastic equilibrium link flows on signal settings for the ENDP, Lee and Machemehl (1998) applied GA optimization heuristic to individual signalized intersection. Their objective function chosen was only a function of green split and the UE link flows. They reported that the results of the GA optimization heuristic could be compared with the local search method due to Sheffi and Powell (1983). Yin (2000) applied the GA notion to a small-scale example network to maximise the network reserve capacity by explicitly taking into account only the green split as a signal timing parameter for an isolated signalized junction. UE assignment was used to obtain equilibrium link flows resulting from the upper level problem. The algorithm performance and convergence were compared with the sensitivity-analysis-based (SAB) algorithm due to Yang and Yagar (1995). The algorithm showed different convergence behaviors, but the resultant solutions were about the same as those of the SAB algorithm in their small-scale example network. The offsets and common network common cycle time optimization as a whole were not taken into account by Yang and Yagar (1995) and Yin (2000) and under various levels of origin–destination (OD) demand increase.

The optimization models developed so far for the ENDP are either calculus based and mathematically lengthy or are of the TRANSYT-Hill-Climbing variety. Both approaches are prone to local optima. Hence, there is a need for a new optimization method which takes into account the interdependency between traffic control and stochastic traffic assignment for various levels of traffic loads. By integrating the genetic algorithms, traffic assignment and traffic control, the GATRANSPFE (Genetic Algorithm, TRANSYT and the PFE) can solve the ENDP problem. The upper-level problem minimises total network performance index as in TRANSYT (Robertson, 1969; Vincent et al., 1980), for which signal timing parameters are generated and feed into the lower-level problem. The lower-level problem is formulated as a path-based SUE to find the stochastic equilibrium link flows using the path flow estimator (Bell et al., 1997).

The deficiency of current solution methods of the ENDP is that there are no consideration for the appropriate expressions to include the offsets between junctions and the network common cycle time in the travel time function as a total system performance index (PI) when level of demand are increased at a critical stage. It would be better to include the offsets and the network common cycle time in the link cost function during the solution process. But it is difficult to formulate and solve the offsets between junctions and network common cycle time in conventional analytical techniques due to its non-convex nature of the problem. The formulation of offsets is also difficult in conventional mathematical techniques as close-convex set and it is mathematically very demanding (see for details, Chiou, 2003). It would also be better to consider the stochastic variations of the route choice when the signalized junctions are overloaded. In this paper, however, offsets, network common cycle time as well as the stage green time are taken into at the upper-level for the ENDP and solved with the GATRANSPFE under the SUE link flows that are obtained using the path flow estimator (PFE) when level of demand are increased at a pre-specified values.

This paper has been organized in a following way. In Section 2, the basic notations are defined. Section 3 is about the problem formulation. Section 4 is on solution methods for the combined traffic assignment and traffic control problem. Numerical analysis is carried out in Section 5. Conclusions are drawn in Section 6.

2. Notations

PI the total network performance index
D_a the delay in vehicles-hours/hour
S_a the number of vehicle-stops per hour
w_a delay weighting factors per unit time
k_a stop weighting factors per unit time
$\boldsymbol{N} = \{1, 2, 3, \ldots, N_j\}$ N_j nodes each of which represents one of N_j fixed-time signal-controlled junctions
$\boldsymbol{L}$ set of N_L links, that is, $\boldsymbol{L} = \{1, 2, 3, \ldots, N_L\}$ where each traffic stream approaching any junction is represented by its own links
c common cycle time
$c_{\min}$ and $c_{\max}$ minimum and maximum specified cycle time respectively
$\boldsymbol{\theta}$ vector of feasible range of offset variables
$\boldsymbol{\phi}$ vector of duration of green times
I_i intergreen time between signal stages
$\phi_{\min}$ and $\phi_{\max}$ minimum and maximum acceptable duration of the stage green timings
$\mathbf{q} = [q_a; \forall a \in \boldsymbol{L}]$ vector of the average flow q_a on link a
s_a saturation flow on link a
$\boldsymbol{W} = w = (r, s); \forall r \in \overline{\mathbf{R}}, \forall s \in \overline{\mathbf{S}}$ set of origin–destination pairs
$\boldsymbol{P}_w$ set of paths each origin–destination pair w, $\forall w \in W$
$\mathbf{t} = [t_w; \forall w \in \boldsymbol{W}]$ vector of travel demand between each origin–destination pairs
$\mathbf{h} = [h_p; \forall p \in \boldsymbol{P}_w, \forall w \in \boldsymbol{W}]$ vector of all path flows, where element h_p is traffic flow on path p

$\boldsymbol{\delta} = [\delta_{ap}; \forall a \in \boldsymbol{L}, \forall p \in \boldsymbol{P}_w, \forall w \in \boldsymbol{W}]$ the link/path incidence matrix where $\delta_{ap} = 1$ if link a is on path p, and $\delta_{ap} = 0$ otherwise
$\boldsymbol{\Lambda}$ OD-path incidence matrix
$\mathbf{y} = [y_w; \forall w \in \boldsymbol{W}]$ the expected minimum origin–destination cost and summation is over all links
$\mathbf{g}(\mathbf{q}, \boldsymbol{\psi}) = [g_p; \forall p \in \boldsymbol{P}_w, \forall w \in \boldsymbol{W}]$ vector of path travel times
$\mathbf{c}^0 = [c_a^0; \forall a \in \boldsymbol{L}]$ vector of free-flow link travel times, where element c_a^0 is the free-flow travel time for link a
$\mathbf{c}(\mathbf{q}, \boldsymbol{\psi}) = [c_a(q_a, \boldsymbol{\psi})]$ vector of all link travel times, where element $c_a(q_a, \boldsymbol{\psi})$ is travel time on link a as a function of flow on the link itself and the signal setting variables
$\mathbf{K} = [K_p^w; \forall p \in \boldsymbol{P}_w, \forall w \in \boldsymbol{W}]$ matrix of link choice probabilities
m number of stages for each signalized junction
M total number of stages at road network
k the number of signal timing variables on a whole road network, the dimension of the problem is $k = \sum_{i=1}^{N} m_i + N$
l_i number of binary bits for each signal timing
$\mathbf{x}$ vector of chromosomes
$\boldsymbol{\psi}$ whole vector of signal timings
$\boldsymbol{\psi}_{\min}$ whole lower bound vector of signal timings
$\boldsymbol{\psi}_{\max}$ whole upper bound vector of signal timings

The following conservation relationships encountered in this paper are the following:

$$\mathbf{q} = \boldsymbol{\delta}\mathbf{h}; \quad \mathbf{t} = \boldsymbol{\Lambda}\mathbf{h}; \quad \mathbf{q} = \mathbf{K}\mathbf{t}$$

3. Formulations

The bi-level formulation of the ENDP is presented as an interaction between decision-takers (travellers) and a decision-maker (signal control) in the following way.

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

$$\text{subject to} \quad c_{\min} \leqslant c \leqslant c_{\max} \tag{2a}$$

$$0 \leqslant \theta \leqslant c \tag{2b}$$

$$\phi_{\min} \leqslant \phi \leqslant \phi_{\max} \tag{2c}$$

$$\sum_{i=1}^{m} (\phi_i + I_i) = c \tag{2d}$$

$\mathbf{q}^*(\boldsymbol{\psi})$ is implicitly obtained by solving by

$$\underset{\mathbf{q}}{\text{Minimise}} \quad \mathrm{z}(\mathbf{q}, \boldsymbol{\psi}) = -\mathbf{t}^{\mathrm{T}}\mathbf{y}(\mathbf{q}, \boldsymbol{\psi}) + \mathbf{q}^{\mathrm{T}}\mathbf{c}(\mathbf{q}, \boldsymbol{\psi}) - \sum_{\mathrm{a} \in \boldsymbol{L}} \int_0^{\mathrm{q_a}(\boldsymbol{\psi})} \mathrm{c_a}(\boldsymbol{\psi}, \mathrm{x})\mathrm{dx}$$
$$\text{subject to} \quad \mathbf{t} = \boldsymbol{\Lambda}\mathbf{h}, \quad \mathbf{q} = \boldsymbol{\delta}\mathbf{h}, \quad \mathbf{h} \geqslant \mathbf{0} \tag{3}$$

where all the notation is previously stated, due to Fisk (1980), leads to a logit path choice model. Following Bell and Iida (1997);

At equilibrium, this function has value zero

$$z(\mathbf{q}^*, \boldsymbol{\psi}) = \mathbf{0} \tag{4}$$

It is possible to infer how equilibrium link flows change with signal setting parameters. Note that

$$\mathrm{d}\mathbf{q}/\mathrm{d}\boldsymbol{\psi} = (\partial z(\mathbf{q}^*, \boldsymbol{\psi})/\partial \mathbf{q})^{-1}(\partial z(\mathbf{q}^*, \boldsymbol{\psi})/\partial \boldsymbol{\psi}) \tag{5}$$

provided that the matrix $(\partial z(\mathbf{q}^*, \boldsymbol{\psi})/\partial \mathbf{q})$ is invertible. Regarding the right hand side of (5)

$$(\partial z(\mathbf{q}^*, \boldsymbol{\psi})/\partial \boldsymbol{\psi}) = \mathbf{I} - \sum_{\text{all } j} t_j - (\partial \mathbf{k}_j/\partial \mathbf{c})\mathbf{J} \tag{6}$$

where $\mathbf{k}_j$ is the jth column of the matrix of link choice proportions $\mathbf{K}$, $\mathbf{I}$ is the identity matrix and the $\mathbf{J}$ is the *Jacobian* of the link cost functions.

In the case of logit model as in PFE with deterrence parameter it may be readily verified that

$$\partial k_{ij}/\partial c_{i'} = \begin{cases} -\alpha k_{ij} + \alpha k_{ij}^2 & \text{if } i = i' \\ -\alpha k_{ii'j} + \alpha k_{ij}k_{i'j} & \text{otherwise} \end{cases}$$

where $k_{ii'j}$ is the proportion of traffic from jth trip table element using both links i and i'.

Concerning the second term on the right hand side of (5),

$$(\partial z(\mathbf{q}^*, \boldsymbol{\psi})/\partial \boldsymbol{\psi}) = -\sum_{\text{all } j} t_j(\partial \mathbf{k}_j/\partial \mathbf{c})(\partial \mathbf{c}/\partial \boldsymbol{\psi}) \tag{7}$$

where $k_{ii'j}$ is the proportion of traffic from jth trip table element using both links i and i'.

Concerning the second term on the right hand side of (5),

$$(\partial z(\mathbf{q}^*, \boldsymbol{\psi})/\partial\psi) = -\sum_{\text{all } j} t_j(\partial\mathbf{k}_j/\partial\mathbf{c}) - \mathbf{K}(\partial\mathbf{t}/\partial\mathbf{z})\mathbf{K}^{\mathrm{T}}(\partial\mathbf{c}/\partial\boldsymbol{\psi}) \tag{8}$$

The evaluation of both k_{ij} and $k_{ii'j}$ can be obtained in Bell and Iida (1997).

Mutually consistent formulation of the combined traffic assignment and area traffic control problem is as follows: Find ψ^* such that

$$\begin{aligned} &\underset{\boldsymbol{\psi}\in\boldsymbol{\Omega}_0}{\text{Minimise}} \quad \mathrm{PI}(\boldsymbol{\psi}, \mathbf{q}) \\ &\text{subject to} \quad \mathbf{q} = \bar{\mathbf{q}} \end{aligned} \tag{9}$$

and

$$\begin{aligned} &\underset{\mathbf{q}\in\mathbf{S}_0}{\text{Minimise}} \quad \mathrm{z}(\boldsymbol{\psi}, \mathbf{q}) \\ &\text{subject to} \quad \boldsymbol{\psi} = \overline{\boldsymbol{\psi}} \end{aligned} \tag{10}$$

where $\bar{\mathbf{q}}$, $\overline{\boldsymbol{\psi}}$ are respectively the fixed values of equilibrium link flows and signal settings, and $\mathbf{S}_0$ is the set of stochastic user equilibrium flows.

GA formulation of the bi-level problem the combined traffic assignment and area traffic control problem is as follows:

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

Decision variables	$\boldsymbol{\psi}$	$=$	$\vert c$	$\vert\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_n, \ldots, \boldsymbol{\theta}_N\vert$	$\vert\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_n, \ldots, \boldsymbol{\phi}_M\vert$
Mapping	$\downarrow$		$\downarrow$	$\downarrow$	
Chromosome (string)	$\mathbf{x}$	$=$	01010101|010101	............010111|10101	............00101101|

The length of the sub-strings (i.e. the number of binary bits per decision variable) can be determined with the desired precision of decision variables by the following relationships (Goldberg, 1989):

$$\beta \geqslant \log_2\left(\frac{(\boldsymbol{\psi}_{\max} - \boldsymbol{\psi}_{\min})}{\pi} + 1\right) \tag{11}$$

where β denotes the length of sub-string (i.e. required number of bits per variable); π denotes the desired precision of variables. On the other hand, the precision of increment rate per decision variable may be calculated as follows:

$$\pi_i = \frac{\boldsymbol{\psi}_{\max} - \boldsymbol{\psi}_{\min}}{2^{l_i} - 1} \quad i = 1, 2, \ldots, k \tag{12}$$

Then, the mapping from a binary string into a real numbers ψ from the range $[\boldsymbol{\psi}_{\min}, \boldsymbol{\psi}_{\max}]$ is carried out using following relation:

$$\psi_i = \psi_{i,\min} + \Phi_i \frac{\psi_{i,\max} - \psi_{i,\min}}{2^{l_i} - 1} \quad i = 1, 2, 3, \ldots, k \tag{13}$$

where Φ_i is the integer resulting from binary representation of the signal timing variable.

Each chromosome, x_i, as a potential solution is represented by a binary string of length $l = \sum_{i=1}^{k} l_i$; the first l_1 bits map into a value from the range $[c_{\max}, c_{\min}]$, the next group of l_2 bits map into a value from the range $[0, c]$, and the last group of l_k bits map into a value from the range $[\phi_{\min,n}, \phi_{\max,n}]$ and so on. The detailed decomposition for the signal timing formula can be obtained in Ceylan and Bell (2004) and Ceylan (2002).

Each signal timing is decomposed in the following way:

(a) Common network cycle time

$$c = c_{\min} + \Phi_i \frac{(c_{\max} - c_{\min})}{2^{l_i} - 1} \quad i = 1 \tag{14}$$

(b) For offsets

$$\boldsymbol{\theta} = \boldsymbol{\Phi}_i \frac{c}{2^{l_i} - 1} \quad i = 2, 3 \ldots, N \tag{15}$$

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

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

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

(c) For stage green timings

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

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

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

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:

$$\phi_i = \phi_{\min,i} + \frac{p_i}{\sum_{k=1}^{m} p_i} \left(c - \sum_{k=1}^{m} I_k - \sum_{k=1}^{m} \phi_{\min,k} \right) \quad i = 1, 2, \ldots, m \tag{16}$$

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 can be obtained in Ceylan and Bell (2004).

4. Solution methods

4.1. The MC solution for the ENDP

The MC process was originally proposed by Allsop (1974) and Gartner (1974) also carried out by Cantarella et al. (1991). 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 UE assignment for the resulting signal timings 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 traffic assignment was a deterministic equilibrium assignment. 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 mutually consistent calculation is carried out using the PFE/TRANSYT interface. It is an iterative optimization and assignment procedure that is very similar in nature to other iterative procedures in the literature. The problem of signal optimization with fixed flow pattern is solved by the TRANSYT program whilst the PFE program solves the SUE assignment with fixed signal settings in the following form:

Step 0. Set $k = 0$ for given signal timings $\boldsymbol{\psi}^{(k)}$, find the corresponding equilibrium flows $\mathbf{q}^*(\boldsymbol{\psi}^{(k)})$ by way of the PFE.

Step 1. Run TRANSYT program to obtain the TRANSYT-optimal signal timings $\boldsymbol{\psi}^{(k+1)}$ for the flows $\mathbf{q}^*(\boldsymbol{\psi}^{(k)})$.

Step 2. Update the travel time function of all links to obtain

$$\boldsymbol{c}\left(\mathbf{q}^*, \boldsymbol{\psi}^{(k+1)}\right) = \boldsymbol{c}^0 + \boldsymbol{d}\left(\mathbf{q}^*, \boldsymbol{\psi}^{(k+1)}\right)$$

Step 3. Calculate the corresponding equilibrium flows $\mathbf{q}^*(\boldsymbol{\psi}^{(k+1)})$ by way of the PFE.

At Step 3, method of successive averages (MSA) (Sheffi, 1985) smoothing is applied to the equilibrium link flows and turning flows in the iterative process in order to overcome fluctuations on equilibrium link flows. The application of the MSA smoothing to flows $\mathbf{q}^*(\boldsymbol{\psi}^{(k)})$ is carried out using the following relationship.

$$q_a^{(k+1)} = \frac{1}{k} q_a^{(k-1)} + \left(1 - \frac{1}{k}\right) q_a^{(k)}$$

Step 4. Run TRANSYT program again to obtain the optimal signal timings $\boldsymbol{\psi}^{(k+2)}$ given by the $\mathbf{q}^*(\boldsymbol{\psi}^{(k+1)})$.

Step 5. Compare the values of $\boldsymbol{\psi}^{(k+2)}$ and $\boldsymbol{\psi}^{(k+1)}$, if there is no change between $\boldsymbol{\psi}^{(k+1)}$ and $\boldsymbol{\psi}^{(k+2)}$ then go to Step 6; otherwise, $k = k + 1$ and go to Step 2.

Step 6. *Stop*: $\boldsymbol{\psi}^{(k+1)}$ and $\mathbf{q}^*(\boldsymbol{\psi}^{(k+1)})$ are the mutually consistent signal timings and equilibrium flows.

The MC process in area traffic control and SUE equilibrium assignment is carried out by sequentially solving the TRANSYT and the PFE. During each iteration, the MSA smoothing process is applied to the SUE flows in order to refrain from premature convergence. The application of the MSA smoothing to the stochastic equilibrium link flows $\mathbf{q}^*(\boldsymbol{\psi}^{(k)})$ is carried out using the following relationship.

$$q_a^{(k+1)} = \frac{1}{k} q_a^{(k-1)} + \left(1 - \frac{1}{k}\right) q_a^{(k)}$$

where k is the iteration number and a is a set of links in $\boldsymbol{L}$.

The stopping criterion for the MC process and the corresponding equilibrium flows will be met when the difference in values of signal timings or of SUE flows between successive iterations is not greater than a predetermined threshold value.

4.2. GATRANSPFE solution for the ENDP

GAs are search and optimization procedures motivated by natural selection and genetics. They combine survival of the fittest among string structures to form a search algorithm. In every *generation*, a new set of artificial strings are created based on *fitness*. Although they are random search techniques, they exploit the historical information to find a new search point with expected improved performance. Because of its simplicity, minimal problem restrictions and minimal assumptions on search space, global approach, and implicit parallelism, GAs have been applied to a wide range of problem domains including engineering, sciences, and commerce.

Genetic algorithms differ from conventional search techniques, in that they start with an initial set of random solutions called a *population (pz)*. Each individual in the population is called a *chromosome*, representing a solution to the problem at hand. A chromosome is a string of symbols, usually, but not necessarily, a binary bit string. The chromosomes evolve through successive iterations, called *generations.* According to some measure of fitness, the chromosomes are *evaluated* during each generation. To create the next generation, offspring or children are formed either by: (a) modifying a chromosome using a mutation operator; or (b) merging the chromosomes from current generation using a crossover operator.

Forming the new generation is carried out by (a) selecting, based on fitness values, some of the parents and offspring, and (b) rejecting the others so as to keep population size constant. The algorithm converges to an optimum or near-optimum value of objective function after several generations.

The GATRANSPFE starts with populating the signal timing, and then by integrating TRANSYT and PFE produces the network performance index until pre-specified generation number is completed as can be seen in Fig. 1.

The assignment of the decoded genetic strings to the signal timings can be obtained in Ceylan and Bell (2004). The solution algorithm of the GATRANSPFE is outlined as follows:

Step 0. Initialization. Set the *user-specified GA parameters*; represent the decision variables $\boldsymbol{\psi}$ as binary strings to form a chromosome $\mathbf{x}$ by giving the minimum $\boldsymbol{\psi}_{\min}$ and maximum $\boldsymbol{\psi}_{\max}$ specified lengths for decision variables.

Step 1. Generate the initial random population of signal timings $\mathbf{X}_t$; set $t = 1$.

During the initialization process, the population of chromosomes are created, where each chromosome is a binary vector of required number of binary bits per decision variable. All binary bits for each chromosome are initialized randomly using the random number generator. Due to the given minimum and maximum bounds for the signal timing variables as an input to the GATRANSPFE, the generated sequence for signal timings are not likely to produce infeasible

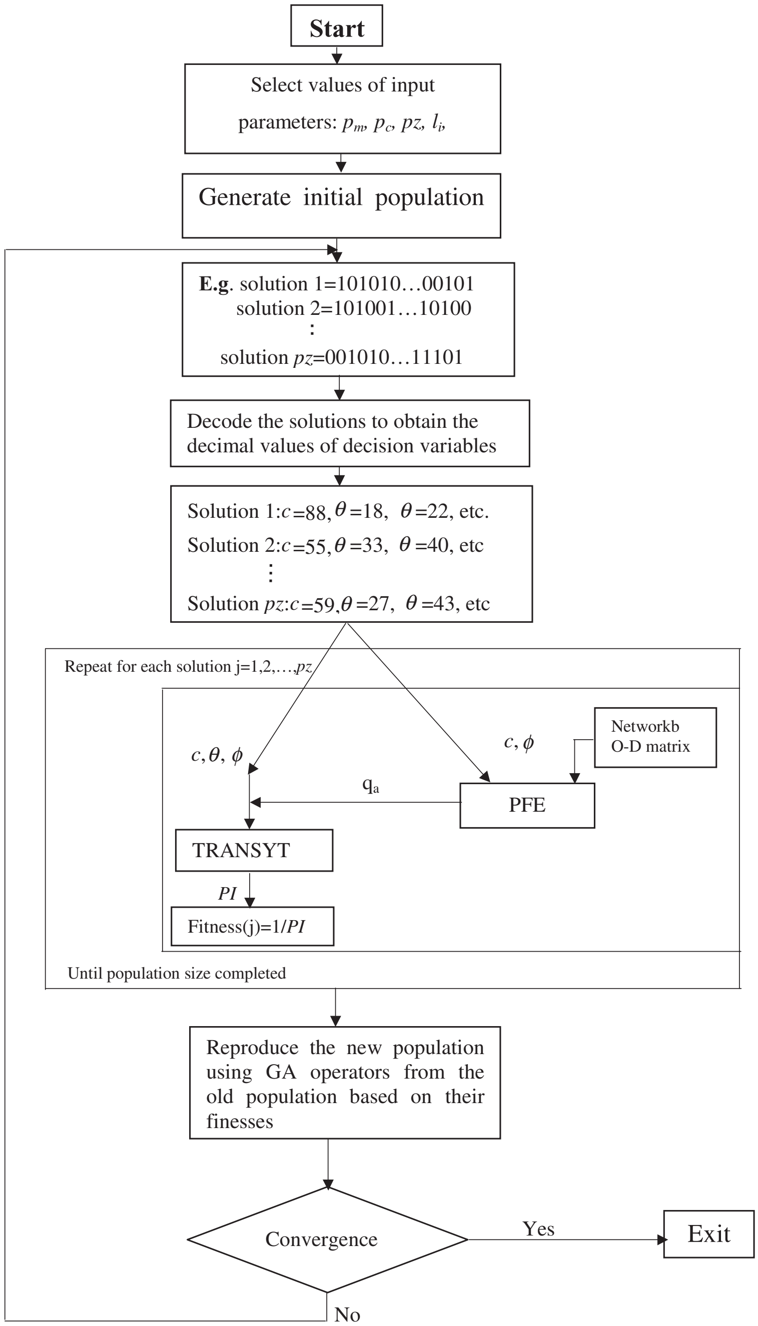

Fig. 1. The flowchart of the GATRANSPFE.

sets. If signal-timing constraints do not satisfy for generated signal settings, the GATRANPSPFE will automatically discards those generated signal timings.

Step 2. Decode all signal timing parameters of $\mathbf{X}_t$ by using (14)–(16) to map the chromosomes to the corresponding real numbers.

Step 3. Solve the lower level problem by way of the PFE. This gives an equilibrium 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) and Chiou (1998), as follows:

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

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 $F(x_j) = \frac{1}{\mathrm{PI}(\psi, q(\psi))}$.

Step 6. Reproduce the population $\mathbf{X}_t$ according to the distribution of the fitness function values.

Step 7. Carry out the crossover operator by a random choice with probability p_c.

Crossover probability (denoted by p_c) is defined as the ratio of the number of offspring produced in each generation to the population size. This ratio controls the expected number $p_c * pz$ of chromosomes to undergo the crossover operation. A higher crossover rate allows exploration of more of the solution space and reduces the chances of settling for a bad local optimum, but the higher the crossover rate, the longer the computation time. After the new population has been filled with crossed over members, mutation can take place. Based on previous studies Goldberg (1989) and Carroll (1996) set the probability of crossover (p_c) between 0.5 and 0.8. Hence, p_c is selected as 0.5 in this study.

Step 8. Carry out the mutation operator by a random choice with probability p_m, then we have a new population $\mathbf{X}_{t+1}$.

Mutation probability (denoted by p_m) is a parameter that controls the probability with which a given string position alters its value. The p_m controls the rate at which new genes are introduced into the population for trial; if it is too low, many genes that would have been useful are never tried out; but if it is too high, there will be much random perturbation, the offspring will start

Table 1
Origin–destination demand for Allsop and Charlesworth's test network in vehicles/hours

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*	100	800
E	300	130	30*	–	20	480
G	550	450	170	60	20	1250
Destination totals	1290	1100	1100	270	1240	5000

* 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.

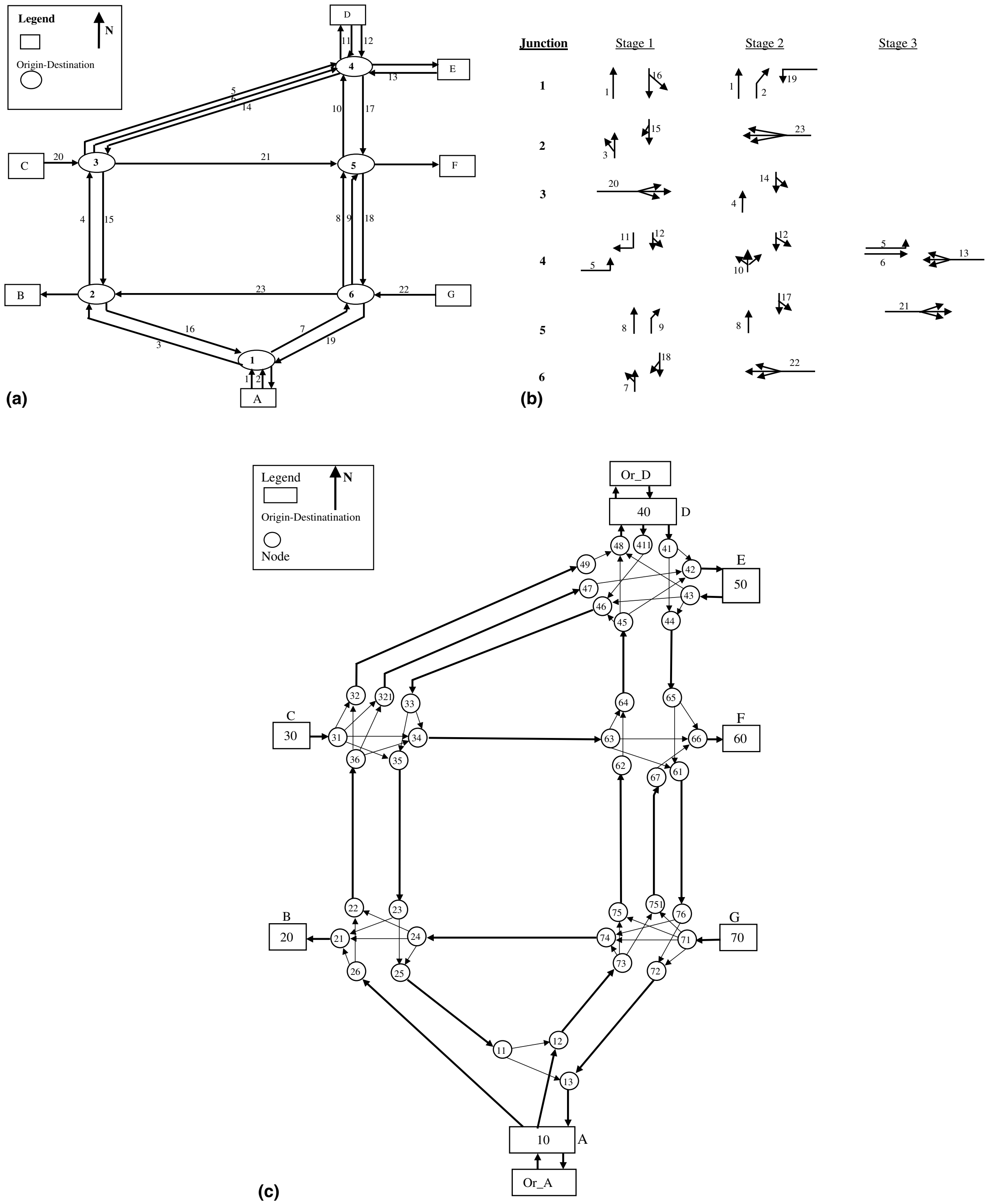

Fig. 2. (a) Layout for Allsop and Charlesworth's test network. (b) Stage configurations for the test network. (c) Representation for traffic assignment use of nodes and links for Allsop and Charlesworth's network.

Table 2
The final values of signal timings resulting from GATRANSPFE when the demand is at 120%

Performance index		Cycle time c (s)	Junction number n	Start of green in seconds		
£/h	veh $-h/h$			Stage 1 $S_{n,1}$	Stage 2 $S_{n,2}$	Stage 3 $S_{n,3}$
			1	21	57	–
			2	63	19	–
1402.8	155.0	96	3	55	20	–
			4	0	35	67
			5	29	50	82
			6	59	1	–

losing their resemblance to the parents, and the algorithm will lose the ability to learn from the history of the search. p_m can be set to $1/pz$ (Carroll, 1996).

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 $t =$ maximal generation number, the chromosome with the highest fitness is adopted as the optimal solution of the problem. Else set $t = t + 1$ and return to Step 2.

5. Numerical example

The application for the GATRANSPFE to the ENDP is tested on an example road network, where the signal timing optimization is carried out at the upper-level based on GA approach, and SUE traffic assignment is solved by means of PFE for various levels of increased demand from the base value given in Table 1 at the lower-level. The test network is illustrated based upon the one used by Allsop and Charlesworth (1977) and Chiou (1998). Basic layouts of the network for use in TRANSYT, in stage configurations and in traffic assignment are given in Fig. 2a–c, where Fig. 2a is adapted from Chiou (1998) and Fig. 2b and c is adapted from Charlesworth (1977). The correspondence links for use in TRANSYT and PFE are set out in Table 2. This numerical test

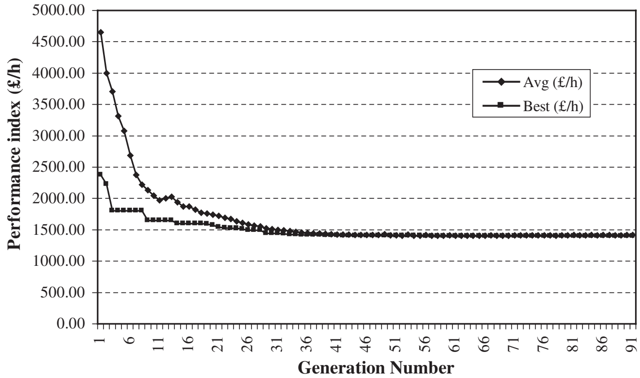

Fig. 3. The GATRANSPFE for Allsop and Charlesworth's test network when demand is at 120%.

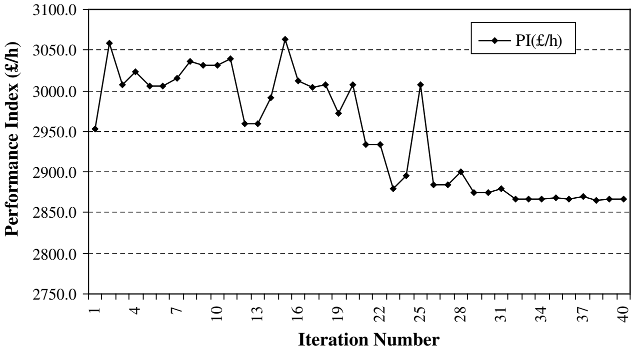

Fig. 4. The MC solution for Allsop and Charlesworth's road network when demand is at 120%.

Table 3
The final values of signal timings resulting from MC when the demand is at 120%

Performance index		Cycle time c (s)	Junction number n	Start of green in seconds		
£/h	veh $-$ h/h			Stage 1 $\Theta_{n,1}$	Stage 2 $\Theta_{n,2}$	Stage 3 $\Theta_{n,3}$
2867.2	330.8	82	1	36	81	0
			2	14	59	0
			3	50	22	0
			4	40	66	10
			5	33	62	8
			6	0	43	0

Table 4
The final values of degree of saturation (%) obtained from the GATRANSPFE when the demand is at 120%

s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	s_{11}	s_{12}	s_{13}
4	59	54	67	68	41	87	61	48	89	93	26	91
s_{14}	s_{15}	s_{16}	s_{17}	s_{18}	s_{19}	s_{20}	s_{21}	s_{22}	s_{23}			
92	73	82	99	70	86	93	98	74	75			

Table 5
The final values of degree of saturation (%) obtained from the MC when the demand is at 120%

s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	s_{11}	s_{12}	s_{13}
4	81	56	77	68	34	116	41	102	83	76	47	75
s_{14}	s_{15}	s_{16}	s_{17}	s_{18}	s_{19}	s_{20}	s_{21}	s_{22}	s_{23}			
83	96	94	133	56	93	91	98	98	95			

includes 20 origin–destination pairs, 23 TRANSYT links, 75 PFE links, 43 feasible routes and 21 signal setting variables at six signal-controlled junction.

The GATRANSPFE is performed with the following user-specified parameters: population size (pz) is 40, reproduction operator is binary tournament selection, crossover probability (p_c) is 0.5, mutation probability (p_m) is 0.025, and the maximal number of generation is 90.

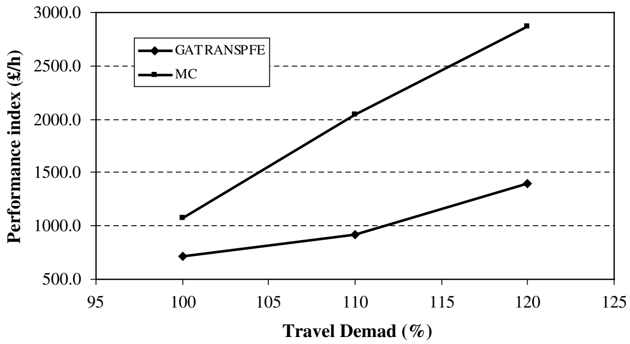

Fig. 5. The comparison of the GATRANSPFE and the MC for the test network when demand is increased to 110% and 120% from base demand.

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

The convergence behavior of the GATRANSPFE and MC can respectively be seen in Figs. 3 and 4. The final values of network performance index and the signal timings for six signalized junctions are also given in Tables 2 and 3. The final values of degree of saturation resulting from the GATRANSPFE and MC can be seen in Tables 4 and 5 when the demand is at 120%.

The application of the GATRANSPFE and the MC method of the ENDP problem for this example network can be seen in Fig. 5 when the demand is increased at 110% and 120% from the base value (i.e. from Table 1). The GATRANSPFE behaved like an undersaturated network in terms of degrees of saturation (less than 90%) on all links (see Ceylan, 2002) when the O–D demand is increased to 110%. The GATRANSPFE still provides a better system performance index than the MC solution but the degree of saturation on some links approaches at critical stage of 100% when the demand is increased to 120%. As can be seen from Fig. 5, the gap between MC and GATRANSPFE becomes higher when the level of demand increases.

The system performance index for this example network are improved about 100% when the values of PI are compared over the MC process (1402.8£/h versus 2867.2£/h) and the network common also are different in both process.

6. Conclusions

In this paper, the GATRANSPFE has been proposed to solve the signalized road under various levels of demand so that better local optima can be achieved. Following earlier test on an uncongested road network (Chiou, 1998; Ceylan, 2002), we went on to implement the GAT-

RANSPFE on the same network under congested conditions for the two levels of increased demand. The results both are obtained from the GATRANSPFE and MC process are compared in terms of efficiency, convergence and degrees of saturation for all links. The results reported here are mainly for 120% increase on travel demand. The reason for is that more than this level of demand, the degrees of saturation on some links such as link 7, 9, 17 are passed critical stage of 100% in the GATRANSPFE.

The proposed GATRANSPFE to solve the ENDP was implemented successfully with different levels of traffic demand for this example network. The solution methods were applied to congested conditions. Better system performances were obtained in the GATRANSPFE in all level of demands when it is compared with the MC solution of the problem. Results showed that GATRANSPFE results for two levels of demands can be compared favorably with the MC solution.

The GATRANSPFE was compared with the MC and it behaved as in the unsaturated case when the demand is increased at 110%, where all degrees of saturation were less than 90%. When the demand increased to 120%, the GATRANSPFE model still provided better system performance but the degree of saturation on some links approached to a critical stage of 100%. The GATRANSPFE model achieved 100% better system performance than the MC process for this example. In particular, the GATRANSPFE showed relatively less sensitivity to the increasing traffic loads as compared to the MC solution.

Concerning the further extensions of this bi-level formulation of the ENDP, detailed examination in a large network needs to be undertaken for congested road networks.

As for the computation efforts for the GATRANSPFE, performed on celeron 1100, each generation for this numerical example was about 7.4 min of CPU time in Fortran 90. The total computation efforts for complete run of the GATRANSPFE model run was 11.1 h. On the other hand, the computation effort for the MC solution on the same machine was performed for each iteration in less than 11.1 s of CPU time and the complete run did not exceed half an hour on that machine.

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 authors would like to thank the anonymous referees for their constructive and useful comments during the development stage of this 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 6th 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.
Cantarella, G.E., Improta, G., Sforza, A., 1991. Iterative procedure for equilibrium network traffic signal setting. Transportation Research 25A (5), 241–249.
Carroll, D.L., 1996. Genetic algorithms and optimizing chemical oxygen–iodine lasers. In: Wilson, H.B., Batra, R.C., Bert, C.W., Davis, A.M.J., Schapery, R.A., Stewart, D.S., Swinson, F.F. (Eds.), Development in Theoretical and Applied Mechanics, Vol. XVIII, Swinson School of Engineering, University of Alabama.
Ceylan, H., 2002. A genetic algorithm approach to the equilibrium network design problem. Ph.D. Thesis, University of Newcastle upon Tyne, UK.
Ceylan, H., Bell, M.G.H., 2004. Traffic signal timing optimisation based on genetic algorithm approach, including drivers' routing. Transportation Research B 38B (4), 329–342.
Charlesworth, J.A., 1977. The calculation of mutually consistent signal settings and traffic assignment for a signal-controlled road network. In: Proceedings of the 7th 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.
Chiou, S.W., 2003. TRANSYT derivatives for area traffic control optimization with network equilibrium flows. Transportation Research 37B (4), 263–290.
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.
Fisk, C., 1984. Optimal signal controls on congested networks. In: 9th International Symposium on Transportation and Traffic Theory. VNU Science Press, pp. 197–216.
Gartner, N.H., 1974. Area traffic control and network equilibrium methods. In: Floran, M. (Ed.), Traffic Equilibrium Methods. Spriger-Verlag, Berlin, pp. 274–297.
Gershwin, S.B., Tan, H.N., 1979. Hybrid optimisation: optimal static traffic control constrained by drivers' route choice behaviour, Massachusetts Institute of Technology, Laboratory for Information and Decision System Report LIDS-p-870.
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.
Robertson, D.I., 1969. TRANSYT: a traffic network study tool, RRL Report, LR 253, Transport and Road Research Laboratory, Crowthorne.
Sheffi, Y., 1985. Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods. MIT, Prentice-Hall, New Jersey.
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.
Wardrop, J.G., 1952. Some theoretical aspects of road research. Proceedings of the Institution of Civil Engineers (Part II) 1 (2), 325–362.
Yang, H., Yagar, S., 1995. Traffic assignment and signal control in saturated road networks. Transportation Research 29A (2), 125–139.
Yin, Y., 2000. Genetic-algorithm-based approach for bilevel programming models. Journal of Transportation Engineering 126 (2), 115–120.