PERGAMON

Control Engineering Practice 10 (2002) 183–195

CONTROL ENGINEERING PRACTICE

www.elsevier.com/locate/conengprac

A multivariable regulator approach to traffic-responsive network-wide signal control ☆

Christina Diakaki*, Markos Papageorgiou, Kostas Aboudolas

Dynamic Systems and Simulation Laboratory, Technical University of Crete, University Campus, 73100 Chania, Greece

Received 27 October 2000; accepted 9 July 2001

Abstract

The paper presents the design approach, the objectives, the development, the advantages, and some application results of the traffic-responsive urban control (TUC) strategy. Based on a store-and-forward modelling of the urban network traffic and using the linear-quadratic regulator theory, the design of TUC leads to a multivariable regulator for traffic-responsive co-ordinated network-wide signal control that is particularly suitable also for saturated traffic conditions. Simulation investigations demonstrate the efficiency of the proposed approach. Results of TUC's first field implementation and evaluation are also presented. Finally, summarising conclusions are drawn and future work is outlined. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Traffic control; Optimal control; Linear-quadratic regulators

1. Introduction

Despite the long-lasting research and developments worldwide (Diakaki, 1999), urban signal control is still an area susceptible of further significant improvements, particularly under saturated traffic conditions. The usually limited availability of space in the urban centres prevents the extension of the existing infrastructure, and, along with the continuously increasing mobility requirements, urge for solutions that will release the serious congestion problems through the best possible utilisation of the available infrastructure. From the control point of view, this may be translated into the employment of actuated systems that respond automatically to the prevailing traffic conditions. This is the aim of the traffic-responsive urban control (TUC) strategy, which was initially developed (Diakaki, Papageorgiou, & McLean, 1999) as part of an integrated traffic control system for corridor networks within the European Telematics Applications in Transport project TABASCO (Telematics Applications in BAvaria, SCotland, and Others).

TUC has been developed to tackle the problem of traffic-responsive network-wide signal control that, particularly under saturated traffic conditions, still lacks an adequate and efficient solution. In contrast to other proposed methods for urban signal control (see Section 2), the *feedback approach* that is pursued by TUC involves the application of systematic and powerful control design methods. The basic philosophy and the importance of these methods are related to their general applicability to any process that can be described by certain types of mathematical models, regardless of the physical process nature (Papageorgiou, 1998). Moreover, in contrast to other proposed methodologies, the specific store-and-forward modelling approach employed in the design of TUC, permits the use of highly efficient optimisation and control methods with polynomial complexity leading to a straightforward network-wide applicability, easy installation and maintenance, as well as low requirements regarding real-time traffic measurements.

The aim of this paper is to present the design approach, the objectives, the development, and the properties of the TUC strategy. Three alternative versions of the multivariable regulator employed by

☆ An early and shorter version of this paper by the authors, entitled "A multivariable regulator approach to traffic-responsive network-wide signal control" was presented at *2000 IFAC Symposium on Control in Transportation Systems* in Braunschweig, Germany, June 2000.

*Corresponding author. Tel.: +30-821-69324/37383; fax: +30-821-69568.

E-mail address: diac@dssl.tuc.gr (C. Diakaki).

PII: S0967-0661(01)00121-6

TUC are presented and discussed. Simulation investigations for a Glasgow (Scotland) based network under different scenarios of demands and incidents are used to demonstrate the efficiency of the proposed method, while the results of its first field implementation are briefly presented to support the findings of the simulation tests. Summarised conclusions are drawn, and future work is outlined.

2. Urban network traffic control

An urban network comprises several streets that cross at junctions which may or may not be signal controlled. An urban junction consists of a set of *approaches* and a common *crossing area*. An approach is a part of a street consisting of one or more lanes. The approach is leading to the common crossing area of the junction so that the traffic on it has right of way (r.o.w.) simultaneously, and a vehicle joining the back of the corresponding queue can expect to pass the signal at roughly the same time, whichever lane it chooses. The traffic at a junction is divided into *streams*. A stream is the smallest portion of traffic considered, and is formed by all vehicles that cross the junction from the same approach. A *saturation flow* is the average flow crossing the stop line of an approach when the corresponding stream has r.o.w., the upstream demand (or the waiting queue) is sufficiently large, and the downstream links are not filled by queues. Two streams are called *compatible* when they can safely cross the junction simultaneously; otherwise, they are called *incompatible* or *conflicting*. A *signal cycle* is one repetition of the basic series of signal combinations at a junction. Its duration is called *cycle time* or simply *cycle*. A *stage* (or *phase*) is a part of the signal cycle, during which one set of streams has the r.o.w. (Fig. 1). Constant *lost times* of a few seconds are interposed between stages in order to avoid interference between incompatible streams of consecutive stages (Fig. 2).

There are four possible ways to influence traffic conditions via traffic lights (Diakaki, 1999):

- *Stage specification*: For complex junctions involving a large number of streams, the specification of the optimal number and constitution of stages is a non-trivial task that can have a major impact on junction capacity and efficiency. This task is typically tackled off-line.
- *Split*: This is the relative green duration of each stage (as a portion of the cycle time) that should be optimised according to the demand of the involved streams.
- *Cycle time*: Longer cycle times typically increase the capacity of the junction because the proportion of the constant lost times becomes accordingly smaller; on the other hand, longer cycle times may increase vehicle delays in undersaturated junctions due to longer waiting times during the red phase.
- *Offset*: This is the time difference between cycles for successive junctions that may give rise to a "green wave" along an arterial; the specification of offset should ideally take into account the possible existence of vehicle queues.

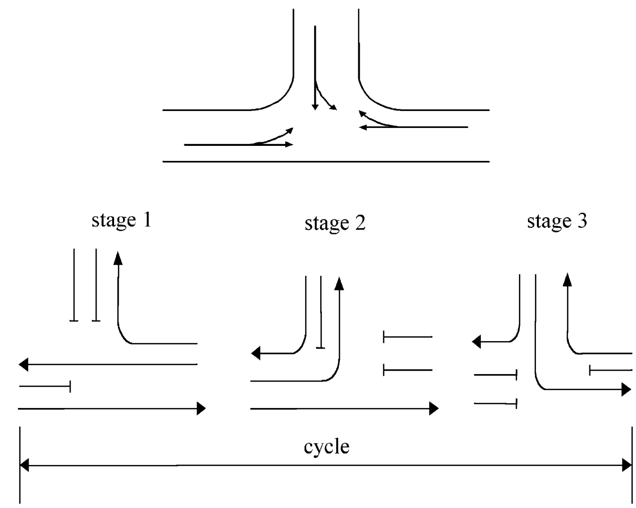


Fig. 1. Example of signal cycle.

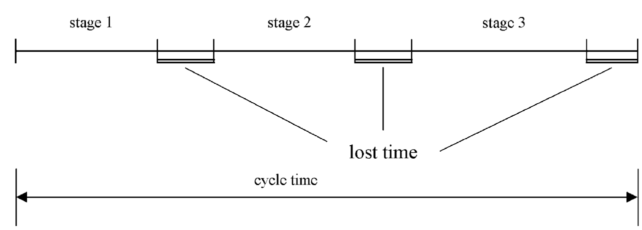


Fig. 2. Cycle time and lost times.

Control strategies employed for road traffic control may be classified according to the following characteristics (Diakaki, 1999):

- *Isolated* strategies are applicable to single junctions, while *co-ordinated* strategies consider an urban area or even a whole network comprising many junctions.
- *Fixed-time* strategies use historical data in order to specify, off-line, optimal time-of-day-dependent plans for the traffic lights; *traffic-responsive* strategies use real-time measurements in order to specify in real time suitable signal settings.
- Some strategies are only applicable to *undersaturated* traffic conditions, whereby vehicle queues are only created during the red phases and are dissolved during the green phases; other strategies are suitable also for *oversaturated* conditions with partially increasing queues that in some cases may reach the upstream junction.

Saturated traffic conditions may cause a serious degradation of the network efficiency for a number of reasons; to start with, a queue that spills back to the upstream junction, causes a waste of green time therein because vehicles heading to the saturated link cannot cross despite having r.o.w.; moreover, vehicles that would turn left or right at the upstream junction are trapped in the corresponding upstream-link queue, thus experiencing delays but also contributing to an accelerated queue growth that may block the further upstream junction and so forth. In some cases, the queue spillbacks may lead to gridlocks in network cycles with a disastrous impact on network efficiency.

The first traffic-responsive urban network control strategies were introduced in the 1980s with the first field implementations of the British SCOOT (Hunt, Robertson, Bretherton, & Royle, 1982), and the Australian SCATS (Lowrie, 1982). SCOOT and SCATS aim at a network-wide co-ordinated control with SCATS adopting a bilevel approach whereby the upper level selects a pre-specified network-wide signal plan while the lower decentralised junction-level modifies (within certain limits set by the upper level) the signal settings so as to respond to the prevailing local traffic conditions. Both SCOOT and SCATS decide on incremental changes of splits, offsets, and cycles based on real-time measurements. For this reason, they have been judged (Dion & Yagar, 1996) to lack a real traffic-responsive behaviour during rapidly changing conditions such as those occurring during the daily business peaks or in case of incidents. Moreover, despite their centralised hardware architecture, both strategies are functionally decentralised with regard to split settings. In other words, the green phase durations at a specific junction depend on real-time measurements from the adjacent upstream approaches only.

More recently, a number of advanced model-based traffic-responsive strategies were developed like OPAC (Gartner, 1983), RHODES (Mirchandani & Head, 1998), PRODYN (Farges, Henry, & Tufal, 1983), CRONOS (Boillot et al., 1992), and UTOPIA (Mauro & Di Taranto, 1989). These strategies do not consider explicitly splits, offsets, and cycles. They formulate the traffic-responsive urban control problem as a combinatorial optimisation problem, and, with the exception of CRONOS, they employ exponential-complexity algorithms to solve for a global minimum. For this reason, these strategies, though conceptually applicable to a whole network, are not real-time feasible for more than one junction, hence they employ heuristic superior control layers with the task of network-wide co-ordination. CRONOS on the other hand, employs a heuristic global optimisation method with polynomial complexity which allows for simultaneous consideration of several junctions, albeit for the price of specifying a local (rather than the global) minimum.

It must be emphasised that most available signal control strategies are not suitable for saturated traffic conditions, which are quite widespread during peak hours in modern cities, because they fail to consider the downstream traffic conditions in their real-time decision-making at individual junctions.

Due to the above reasons, there is lack of efficient, genuinely and systematically co-ordinated control strategies applicable to large-scale networks. Gazis and Potts (1963) had suggested the so-called store-and-forward modelling approach that describes the network traffic flow process so as to circumvent the inclusion of binary variables; this permits the use of highly efficient optimisation and control methods with polynomial complexity for the co-ordinated control of large-scale networks. Based on the store-and-forward modelling approach and the linear-quadratic (LQ) regulator methodology, the proposed TUC approach designs (off-line) and employs (on-line) a multivariable regulator for the traffic-responsive co-ordinated urban network control in a systematic and generic way.

3. The TUC strategy

For the development of a mathematical model, the urban network is represented as a directed graph with links (approaches) $z \in Z$ and junctions $j \in J$. For each signal-controlled junction j, I_j and O_j denote the sets of incoming and outgoing links, respectively. It is assumed that the offset, the cycle time C_j, and the total lost time L_j of junction j are fixed; for simplicity, $C_j = C$ is assumed for all junctions $j \in J$. Furthermore, the signal control of junction j is based on a fixed number of stages that belong to the set F_j, while v_z denotes the set of stages where link z has r.o.w. Finally, the saturation flows S_z, $z \in I_j$, and the turning movement rates $t_{z,w}$, $z \in I_j$, $w \in O_j$, are assumed known and fixed.

By definition, the constraint $\sum_{i \in F_j} g_{j,i} + L_j = C$ holds at junction j, where $g_{j,i}$ is the green time of stage i at junction j. Additionally, $g_{j,i}$ is constrained to lie in the range $[g_{j,i,\min}, g_{j,i,\max}]$ with $g_{j,i,\min}$ and $g_{j,i,\max}$ the corresponding minimum and maximum permissible green times, respectively.

Consider a link (approach) z connecting two junctions M, N such that $z \in O_M$ and $z \in I_N$ (see Fig. 3). The dynamics of link z are expressed by the following equation:

$$x_z(k+1) = x_z(k) + T[q_z(k) - s_z(k) + d_z(k) - u_z(k)], \quad (1)$$

where x_z is the number of vehicles within link z; q_z and u_z are the inflow and outflow, respectively, of link z over the period $[kT, (k+1)T]$ with T the control interval and $k = 1, 2, \ldots$ a discrete time index; and d_z and s_z are the demand and the exit flow, respectively. For the exit flow

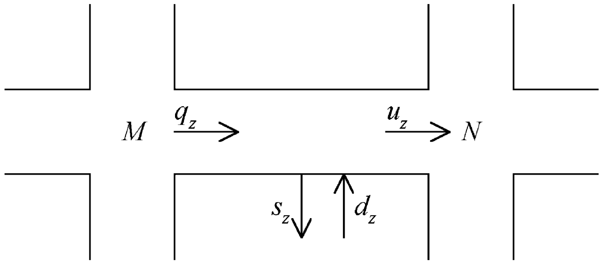


Fig. 3. An urban road link.

the formula $s_z(k) = t_{z,0}q_z(k)$ holds, with exit rates $t_{z,0}$ considered fixed and known.

The inflow to the link z is given by $q_z(k) = \sum_{w \in I_M} t_{w,z}u_w(k)$, where $t_{w,z}$ with $w \in I_M$ are the turning rates towards link z from the links that enter junction M. Assuming that space is available in the downstream links and that x_z is sufficiently high, the outflow u_z of a link is equal to the saturation flow S_z if the link has r.o.w., and equal to zero otherwise. However, if the control interval T is chosen not less than the cycle time C, an average value is obtained by $u_z(k) = S_zG_z(k)/C$, where G_z is the effective green time of link z, calculated as $G_z(k) = \sum_{i \in v_z} g_{j,i}(k)$.

Assume availability of nominal green times $g_{j,i}^{\mathrm{N}}$ that lead to steady-state link queues under non-saturating constant nominal demand d_z^{N}, i.e. we have from (1)

$$(1 - t_{z,0})q_z^N + d_z^N - u_z^N = 0, \tag{2}$$

where q_z^{N}, u_z^{N} are corresponding nominal steady-state values.

Introducing all the above in (1), the following state equation is finally obtained

$$x_z(k+1) = x_z(k) + T\left[(1 - t_{z,0}) \sum_{w \in I_M} \frac{t_{w,z}S_w\left(\sum_{i \in v_w} \Delta g_{M,i}(k)\right)}{C} + \Delta d_z(k) - \frac{S_z\left(\sum_{i \in v_z} \Delta g_{N,i}(k)\right)}{C}\right], \tag{3}$$

where $\Delta g_{j,i} = g_{j,i} - g_{j,i}^{\mathrm{N}}$ and $\Delta d_z = d_z - d_z^{\mathrm{N}}$.

Applying (3) to an arbitrary network comprising several signalised junctions $j \in J$, the following state equation (in vector form) describes the evolution of the system in time

$$\mathbf{x}(k+1) = \mathbf{A}\,\mathbf{x}(k) + \mathbf{B}\,\Delta\mathbf{g}(k) + \mathbf{T}\,\Delta\mathbf{d}(k), \tag{4}$$

where $\mathbf{x}$ is the state vector of the numbers of vehicles x_z within links $z \in Z$; $\Delta\mathbf{g}$ is the vector of $\Delta g_{j,i} = g_{j,i} - g_{j,i}^{\mathrm{N}}$, $\forall i \in F_j, \forall j \in J$; $\Delta\mathbf{d}$ is the vector of $\Delta d_z = d_z - d_z^{\mathrm{N}}$; and $\mathbf{A} = \mathbf{I}, \mathbf{B}$, and $\mathbf{T}$ are the state, input, and disturbance matrices, respectively. Note that the input matrix $\mathbf{B}$ reflects the specific network topology, fixed staging, cycle, saturation flows, and turning rates. In contrast to other applications of store-and-forward modelling (e.g. Gazis, 1964; D'Ans & Gazis, 1976; Singh & Tamura, 1974; etc.), x_z in the above formulation denotes the number of vehicles within a link z instead of the queue length within it, which circumvents the need of incorporating time lags in the state Eq. (4).

In order for the application of the LQ-methodology to lead to a feedback control law without feedforward terms, i.e. a control law that reacts to the manifest impact of the disturbances on the controlled process rather than to disturbance forecasts, $\Delta\mathbf{d}(k) = 0$ is assumed, leading from (4) to the following state equation:

$$\mathbf{x}(k+1) = \mathbf{x}(k) + \mathbf{B}\,\Delta\mathbf{g}(k). \tag{5}$$

It should be stressed here that the LQ optimisation methodology should be viewed as a vehicle for deriving an efficient gain matrix rather than as an attempt to optimise a physical criterion subject to accurate modelling equations and constraints. Within this frame the underlying assumption of zero disturbances is acceptable and (5) is utilised as the mathematical model within the LQ optimal control problem.

In order to minimise the risk of oversaturation and the spillback of link queues, one may attempt to minimise and balance the links' relative occupancies $x_z/x_{z,\max}$, where $x_{z,\max}$ is the storage capacity of link $z \in Z$ (measured in vehicles). A quadratic criterion that considers this control objective has the general form

$$\Im = \frac{1}{2}\sum_{k=0}^{\infty}\left(\|\mathbf{x}(k)\|_{\mathbf{Q}}^2 + \|\Delta\mathbf{g}(k)\|_{\mathbf{R}}^2\right), \tag{6}$$

where $\mathbf{Q}$ and $\mathbf{R}$ are non-negative definite, diagonal weighting matrices. The infinite time horizon in (6) is taken in order to obtain a time-invariant feedback law according to the LQ optimisation theory. The first term in (6) is responsible for minimisation and balancing of the relative occupancies of the network links. To this end, the diagonal elements of $\mathbf{Q}$ are set equal to the inverses of the storage capacities of the corresponding links. Furthermore, the magnitude of the control reactions can be influenced by the choice of the weighting matrix $\mathbf{R} = r\mathbf{I}$. To this end, the choice of r is performed via a trial-and-error procedure so as to achieve a satisfactory control behaviour for a given application network.

Minimisation of the performance criterion (6) subject to (5) leads to a LQ feedback control law

$$\mathbf{g}(k) = \mathbf{g}^{\mathrm{N}} - \mathbf{L}\mathbf{x}(k), \tag{7}$$

where $\mathbf{g}$ is the vector of the green times $g_{j,i}, \forall i \in F_j, \forall j \in J$, and $\mathbf{L}$ is the resulting control matrix, which depends upon the problem matrices $\mathbf{B}$, $\mathbf{Q}$, and $\mathbf{R}$ but was found in simulation investigations to have very low sensitivity with respect to reasonable variations of traffic parameters (such as turning rates, saturation flows, etc.) (Diakaki, 1999). The calculation of $\mathbf{L}$ via solution of the discrete-time Riccati equation may be very time consuming for networks with high dimensions. However, this computational effort is required only off-line, while on-line (i.e. in real-time) the calculations are limited to the execution of (7) with a given constant control matrix $\mathbf{L}$ and state measurements $\mathbf{x}(k)$.

It should be noted that the matrix $\mathbf{L}$ resulting from the formulated LQ problem for specific networks has an overlapping structure (see Section 4 and Fig. 6). Hence, for networks with a very high number (hundreds or thousands) of junctions, heuristic or rigorous overlapping decomposition techniques (Siljak, 1991) may be used to reduce the required off-line computational effort. In any case, even for large-scale networks, the computational effort needed to execute the simple feedback control law of TUC in real time is low, and certainly much lower than that of any other known advanced signal control strategy.

To apply (7), availability of measurements for all state variables is required in real-time. However, the numbers of vehicles x_z are usually not directly measurable, unless video detection systems are utilised. For this reason, local occupancy measurements o_z, collected in real time by traditional detector loops, may be transformed into (approximate) numbers of vehicles via suitable non-linear functions $x_z(k) = f_z(o_z(k))$ (Diakaki, 1999).

Since the LQ-methodology does not take into account the existence of control constraints, the existing constraints are imposed after application of (7). To this end, the following optimisation problem is solved in real-time for each junction j so as to specify feasible green times $G_{j,i}$ that are closest in distance to the non-feasible regulator-based green times $g_{j,i}$ resulting from (7)

$$\min_{G_{j,i}} \sum_{i\in F_j} (g_{j,i} - G_{j,i})^2$$

subject to

$$\sum_{i\in F_j} G_{j,i} + L_j = C$$

$$G_{j,i} \in \left[g_{j,i,\min}, g_{j,i,\max}\right] \forall i \in F_j. \tag{8}$$

This special type of quadratic-programming problem is solved by use of a simple low-cost algorithm that was invented specifically for this problem (Diakaki, 1999). The algorithm reaches the exact solution in a finite number of iterations that does not exceed the number $|F_j|$ of stages at junction j.

Multivariable regulator (7) has a reactive rather than anticipatory behaviour whereby it responds indirectly to unknown disturbances, and therefore it does not need any predictions of the future traffic conditions. However, it should be emphasised that its reactive control behaviour is by no means a myopic one, since it relies on real (measured) state information and is designed on the basis of an infinite optimisation horizon. As mentioned earlier, the sample time interval T of control law (7) cannot be lower than a cycle (e.g. 90 s), which may seem too long compared to other advanced signal control strategies. However, the simulation investigations in Section 4 demonstrate that this sample time length is not an obstacle towards achieving an excellent real-time control behaviour, particularly under saturated traffic conditions.

Structurally, the control matrix $\mathbf{L}$ provides control law (7) with a gating feature so as to protect downstream links from oversaturation. Roughly speaking, the higher the number of vehicles within a particular link z, the more the green times of the links that lead to z are decreased through the application of (7) due to the sign of the corresponding elements of $\mathbf{L}$. In order for the gating feature of the multivariable regulator to be further accentuated, the utilised x_z values may be weighted such that, the higher the value of x_z, the higher its weight. More precisely, the accentuation is achieved if the values of x_z utilised in the control laws are replaced by their weighted counterparts x'_z according to the following relationship

$$x'_z = x_z(k)/[1 - (bx_z(k)/x_{z,\max})] \tag{9}$$

with $b \in [0, 1)$ a parameter that is selected via a trial-and-error procedure so as to achieve a satisfactory accentuation of the gating feature for a given application network.

Control law (7) requires availability of nominal values $\mathbf{g}^{\mathrm{N}}$. Alternatively, control law (7) may be employed in the following form, where $\mathbf{g}^{\mathrm{N}}$ is not needed:

$$\mathbf{g}(k) = \mathbf{g}(k-1) - \mathbf{L}[\mathbf{x}(k) - \mathbf{x}(k-1)]. \tag{10}$$

Control law (10) is obtained by subtracting (7) for control period $k-1$ from (7) for control period k.

A further control law that eliminates the need of nominal values $\mathbf{g}^{\mathrm{N}}$ may be obtained through the formulation of the urban traffic control problem as a linear-quadratic-integral (LQI) optimal control problem based on the same modelling approach and pursuing the same control objective as before. The LQI methodology leads to the following control law (Diakaki, 1999)

$$\mathbf{g}(k) = \mathbf{g}(k-1) - \mathbf{L}_1\mathbf{x}(k) - \mathbf{L}_2\mathbf{x}(k-1) \tag{11}$$

where $\mathbf{L}_1$ and $\mathbf{L}_2$ are control matrices.

After the application of either (10) or (11), the same algorithm mentioned earlier, modifies the calculated green light durations so as to satisfy the constraints. In order to avoid wind-up phenomena in the regulators, the values $\mathbf{g}(k-1)$ required in (9) and (10) are set equal to the *bounded* values of the previous control period (i.e. after application of the constraints).

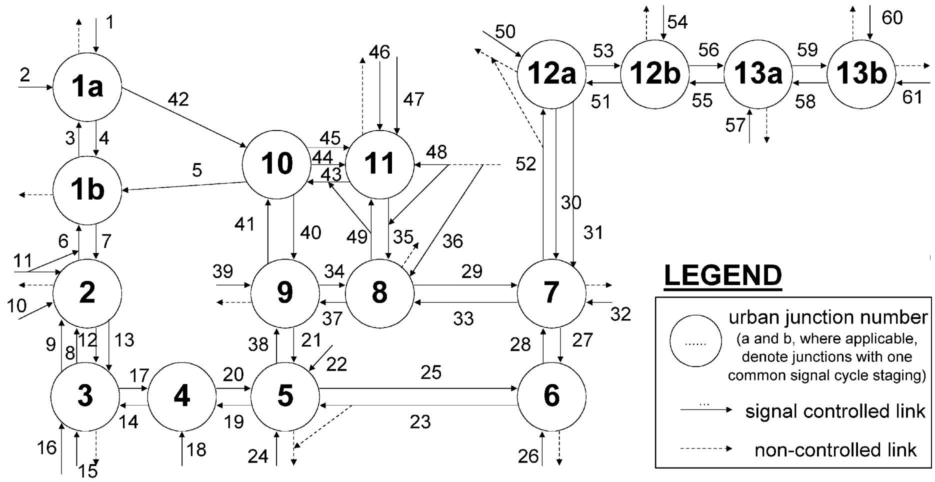


Fig. 4. The example network.

Table 1
Simulation results for fixed-time signal control (FSC) and TUC strategy using accurate x_z-measurements

	Total waiting time		Total travel time		Total time spent		Total fuel consumption	
	veh/h FSC	Percentage change TUC	veh/h FSC	Percentage change TUC	veh/h FSC	Percentage change TUC	veh/l FSC	Percentage change TUC
Scenario (i)	128	−100	2129	−34	2257	−38	3875	−23
Scenario (ii)	2042	−100	3383	−26	5424	−54	5500	−18
Scenario (iii)	4	−100	2365	−19	2369	−20	4237	−13
Scenario (iv)	128	−100	2166	−34	2294	−38	3920	−23
Scenario (v)	120	−100	2108	−32	2228	−35	3849	−21

4. Simulation investigations

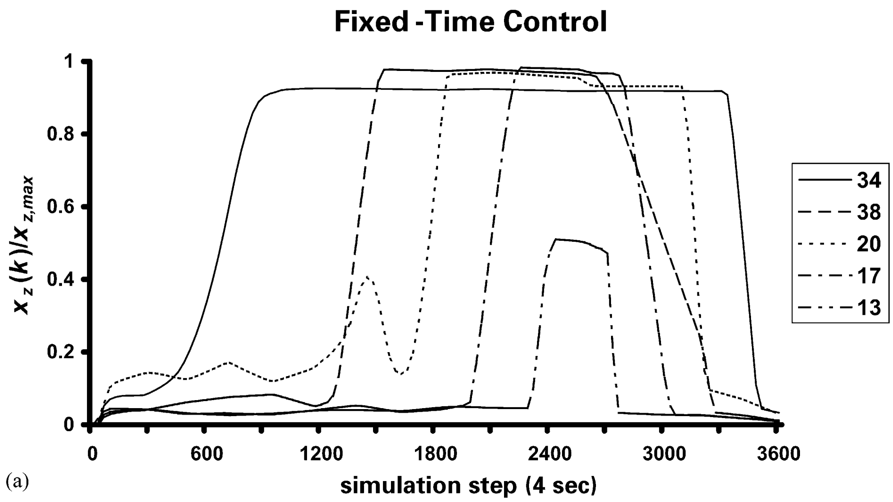
For the simulation investigations of the TUC strategy, the example network of Fig. 4 is used that consists of 13 signalised junctions and 61 links, and is based on the Glasgow network for which the initial development of the TUC strategy took place within the TABASCO project. The investigations are based on 4 h simulations with TUC strategy running every 2 min, a control interval that is equal or twofold to the cycle time of all the considered junctions.

For the simulation tests, the example network is modelled (Diakaki, 1999) via METACOR. METACOR (Elloumi, Haj-Salem, & Papageorgiou, 1994) is a macroscopic modelling tool for simulating traffic flow phenomena within motorway, urban, or corridor (i.e. mixed) networks of arbitrary topology and characteristics. The modelling approach of METACOR allows for the simulation of all kinds of traffic conditions (free, dense, and congested) and of capacity-reducing events (incidents) with prescribed characteristics (location, intensity, and duration). The simulation model of the example network was validated against real data within the TABASCO project for the real Glasgow network (McLean et al., 1997a).

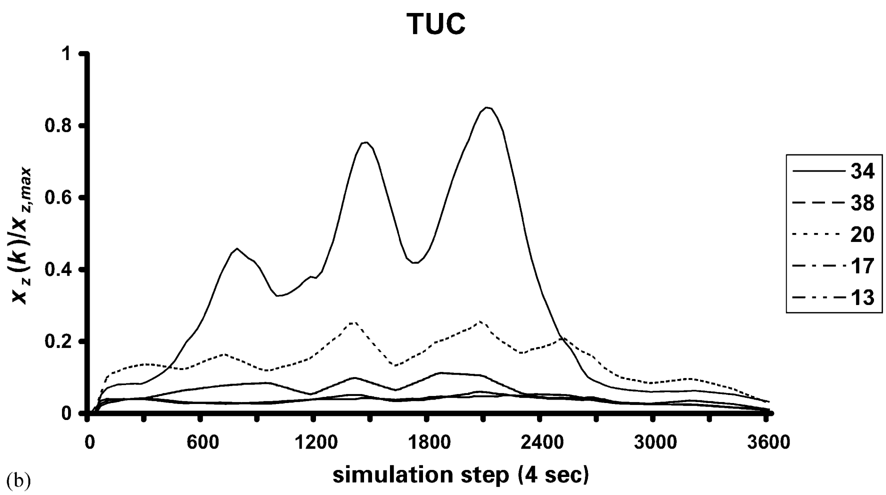
Initially, simulations are performed with field-applied, off-line optimised fixed-time signal control for five different scenarios of demands and incidents. Then, control laws (7), or (10), or (11) are applied and tested with the METACOR simulator for five demand and incident scenarios with the following characteristics:

(i) Low demand in all but a few network origins.
(ii) High demand (approximately 40% higher than scenario (i)) in almost all network origins.
(iii) Demands present high time-fluctuations between the extremes of scenarios (i) and (ii).
(iv) Demands like scenario (i) and a major incident occurring before the peak period.
(v) Demands like scenario (i) and a major incident occurring during the peak period.

Vehicles are pushed into the simulated network at specific origin links according to the particular scenario demand. If, however, a network congestion spills back



(a)



(b)

Fig. 5. Comparison of fixed-time signal control and TUC for scenario (i).

into an origin link, the demand that cannot enter the simulated network, is stored in a virtual waiting queue at that origin, and is pushed into the network as soon as space becomes available on the origin link.

Table 1 summarises the results of the simulation tests in terms of the performance indices:

- *Total waiting time* of all vehicles in all virtual origin queues; note that this performance index reflects, to some extent, the impact of congestion beyond the simulated network part.
- *Total travel time* of all vehicles within the simulated network links.
- *Total time spent* is the sum of both previous criteria.
- *Total fuel consumption*.

The performance indices are calculated by METACOR for the 4 h simulation horizon with TUC employing the multivariable regulator (7). In these investigations, it is assumed that the numbers of vehicles $\mathbf{x}(k)$ within the urban links are accurately measurable in real-time (e.g. through a video detection system) to feed control law (7). Table 1 indicates that the TUC strategy leads to a significant reduction of all performance indices. More specifically, the total waiting time at the network origins is reduced by 100% for all investigated scenarios (i.e. no virtual queues are created), while the total travel time, the total time spent, and the total fuel consumption are reduced in the ranges of 19–34%, 20–54%, and 13–23%, respectively, depending upon the investigated scenario. Note that improvements are higher for higher-saturation scenarios.

Fig. 5 displays the time evolution of the relative occupancies within some selected links under fixed-time signal control and the TUC strategy employing (7), for the demand scenario (i). By inspection of the network sketch in Fig. 4 and the diagrams of Fig. 5, one may see that under fixed-time signal control, congestion develops in link 34 at junction 8. This congestion spills back through links 38, 20, 17, and reaches link 13. Under the TUC strategy the same congestion does not even reach junction 9. Similar performance is also achieved in the other investigated demand scenarios. The application of

TUC largely avoids the development of extended oversaturated conditions and consequently the waste of green time.

If the number of vehicles within urban links utilised in (7) are estimated through local occupancy measurements, the achieved amelioration of the traffic conditions is lower but still significant as compared to the fixed-time signal control. More specifically, the total waiting time at the network origins, the total travel time, the total time spent, and the total fuel consumption are reduced in the respective ranges 73–100%, 10–30%, 14–34%, and 6–20% (Diakaki, 1999).

Since the control matrix **L** resulting from the Riccati-equation has no strictly zero elements, the control of each junction depends on the measurements $x_z(k)$ from *all* network links (co-ordinated network control). Intuitively, however, far-distant links have a negligible impact on the signal settings of a junction. Therefore, after the calculation of matrix **L**, its elements that are found to have a minor effect on the control behaviour (usually corresponding to the most distant links) are nullified. To find the elements of **L** that can be nullified without a significant effect on the control decisions, the following test is performed. The elements of **L** are rounded off at gradually decreasing decimal points, thus leading to the gradual nullification of the low-valued elements, and the control results are examined via simulation. The nullification stops when the resulting control performance starts to deteriorate significantly, as this means that the further nullification would neglect elements of **L** that are important for the control decisions. The control matrix **L** with elements rounded off at 2 decimal points was finally selected, because a non-negligible deterioration appears when the rounding off reaches one decimal point (Diakaki, 1999). Despite this nullification, the resulting gain matrix is not as sparse as one might expect, and the effects of relatively distant links are still present at a higher or lower degree. In fact, the control decisions at each junction depend upon the traffic conditions within the majority of the network links. Fig. 6 shows graphically, for some representative junctions, the network links that affect their signal control via positive or negative values of the corresponding elements of **L**. From Fig. 6, it becomes obvious that the signal control of a junction is affected by the traffic conditions even of relatively distant links. This holds true not only for junctions located centrally within the urban network (Fig. 6b), but even for junctions located at the network boundaries (Fig. 6a and c). It is judged that this central (co-ordinated) way of controlling the network's splits is a major reason for the exceptional efficiency of TUC. Nevertheless, if TUC is to be applied to networks with a very high number of junctions, the resulting control matrix **L** is expected to have an overlapping structure, i.e. control of each junction will be based on measurements from a network area around the junction, not from the whole network.

The simulation investigations of both (10) and (11) using measured or (occupancy-based) estimated values of numbers of vehicles within links indicate a similar performance with the control law (7) (Diakaki, 1999). Given this similar performance, the selection of the approach to be employed may be based on other criteria like e.g. requirement of network authorities to utilise nominal values or lack of nominal values, etc.

5. Real-life implementation and evaluation

The field implementation and evaluation of the TUC strategy took place for a part of the M8 corridor network in Glasgow (McLean et al., 1998) within the TABASCO project. In this field implementation, TUC actually consisted a part of the integrated traffic control strategy IN-TUC (INtegrated-Traffic-responsive Urban Control). IN-TUC (Diakaki, 1999) has been designed so as to integrate three different control applications, namely an urban traffic control strategy, a ramp metering strategy, and a route guidance strategy. For the urban traffic control part of IN-TUC, the TUC strategy is employed. For ramp metering, the ALINEA (Asservissment LINeaire d'Entrée Autoroutière) strategy, a local feedback control law derived by use of classical feedback methods is used (Papageorgiou, Haj-Salem, & Blosseville, 1991). The aim of ALINEA is to regulate the traffic flow entering the motorway so as to maintain the mainstream occupancy downstream of the merging location at a desired pre-specified level. This way, ALINEA manages to reduce motorway congestion and increase the motorway throughput. Finally, for route guidance, availability of variable message signs (VMSs) is assumed that provide suggestions to the drivers on the routes to follow so as to minimise their individual travel times. For this part of the strategy, a simple feedback control law is applied to each VMS with the aim of equalising the travel times between both corresponding alternative routes, which leads to user-optimal conditions (Pavlis & Papageorgiou, 1999). The aforementioned parts of the strategy are integrated in the sense of the mutual exchange of measurements and decisions as Fig. 7 displays. Each part may run with a different control interval and each part may be used as an independent control strategy.

The M8 motorway through Glasgow carries both long-distance traffic within Scotland and local traffic between the eastern and western sectors of the city of Glasgow. During peak hours, local traffic interacting with through traffic at junctions causes major delay to motorway traffic. This situation is most acute at the east-bound on-slip of junction 16 (see Fig. 8 (McLean et al.,

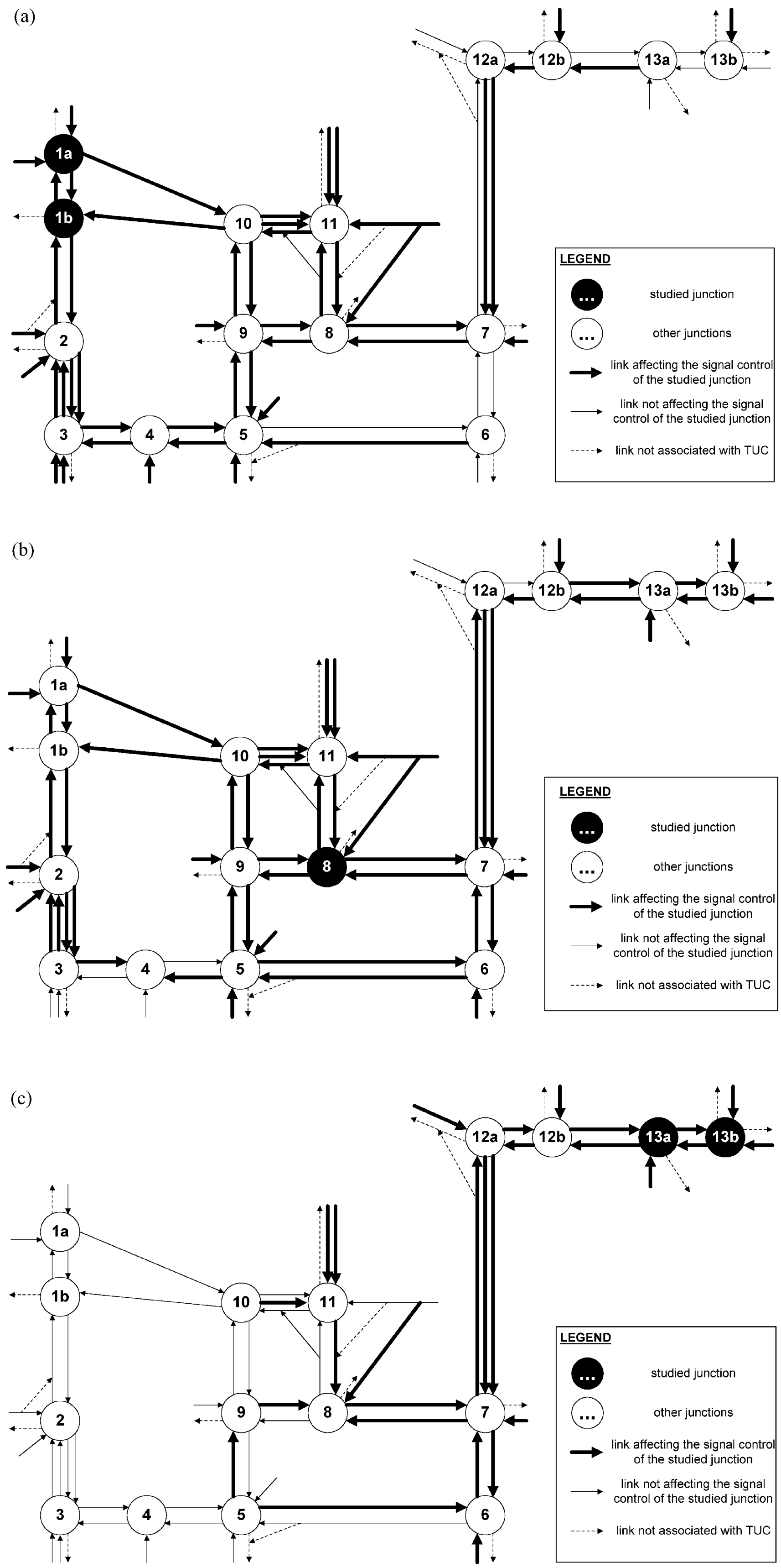


Fig. 6. Network links (bold) that affect signal control of (a) junction 1, (b) junction 8, and (c) junction 13 under LQ control.

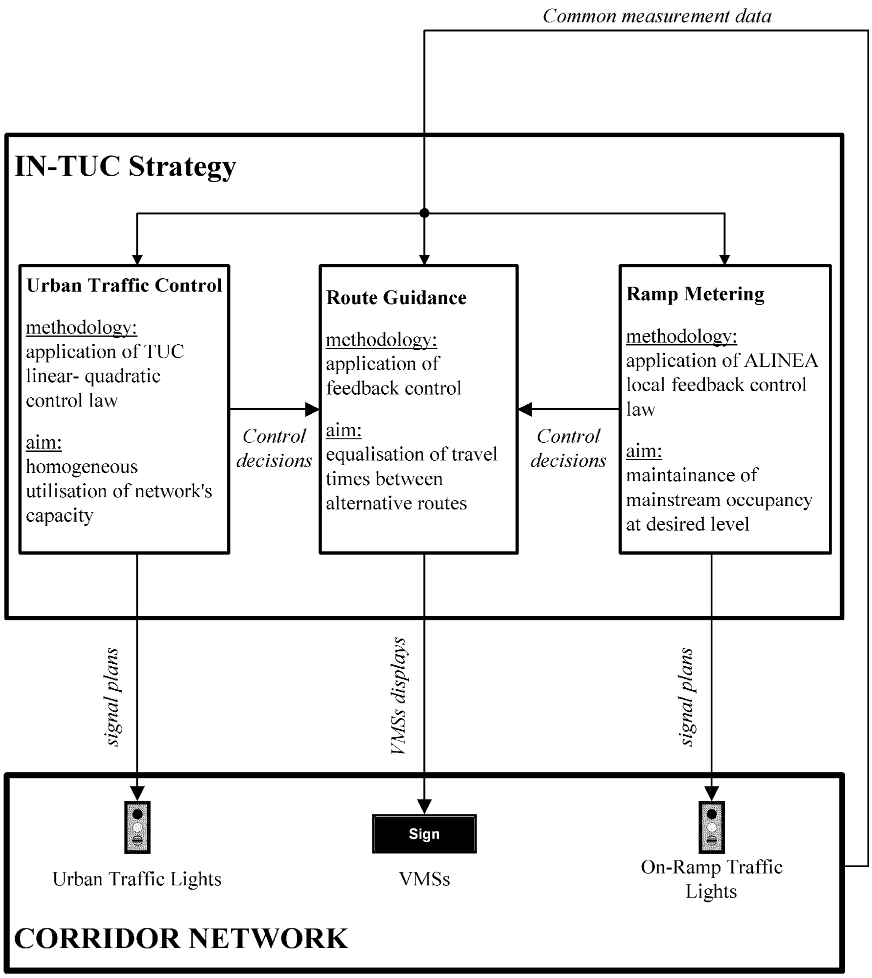


Fig. 7. Functional architecture of IN-TUC strategy.

1997a)). There are several alternative urban routes to the M8, which also suffer severe congestion in the peak hours. These routes are close to the central business district of the city and free passage is hindered by many signalised intersections.

The control elements included in the network area displayed in Fig. 8, which should be used in order to reduce congestion, are ramp metering at Craighall on-ramp (junction 16), several VMSs installed in the urban area around junction 16, and signal control of the urban junctions along the urban routes parallel to the M8. The optimal and integrated utilisation of these control elements was expected, according to past simulation studies (Diakaki, Papageorgiou, & McLean, 1997), to lead to a sensible improvement of the network management and reduction of traffic problems. To this end, the integrated traffic-responsive urban corridor control strategy IN-TUC was applied to the M8 corridor network.

IN-TUC has been applied to the Glasgow network with control intervals of 2 min for the signal control and route guidance, and 60 s for ramp metering (for more details on the structure of the strategy for the Glasgow implementation see McLean et al., 1997b). Regarding the control decisions of TUC, the road authorities specified that the following operational constraint should be respected (McLean, et al., 1997b). The final control output should be a set of plan numbers, where each number corresponds to one out of six pre-specified urban control plans available for every junction. To this end, after the calculation of the green times via TUC, a matching process is followed in order to determine the available plan that best matches the calculated one. This, however, constrains the TUC strategy from optimal operation since the number of available plans is extremely limited compared to all possible outcomes of TUC. Despite this imposed operational constraint, TUC and IN-TUC as a whole demonstrated, as discussed below, a rather high efficiency.

The field implementation and evaluation in Glasgow was phased. It started with the installation of ramp metering with ALINEA in July 1997. TUC was launched in November 1997, while IN-TUC became

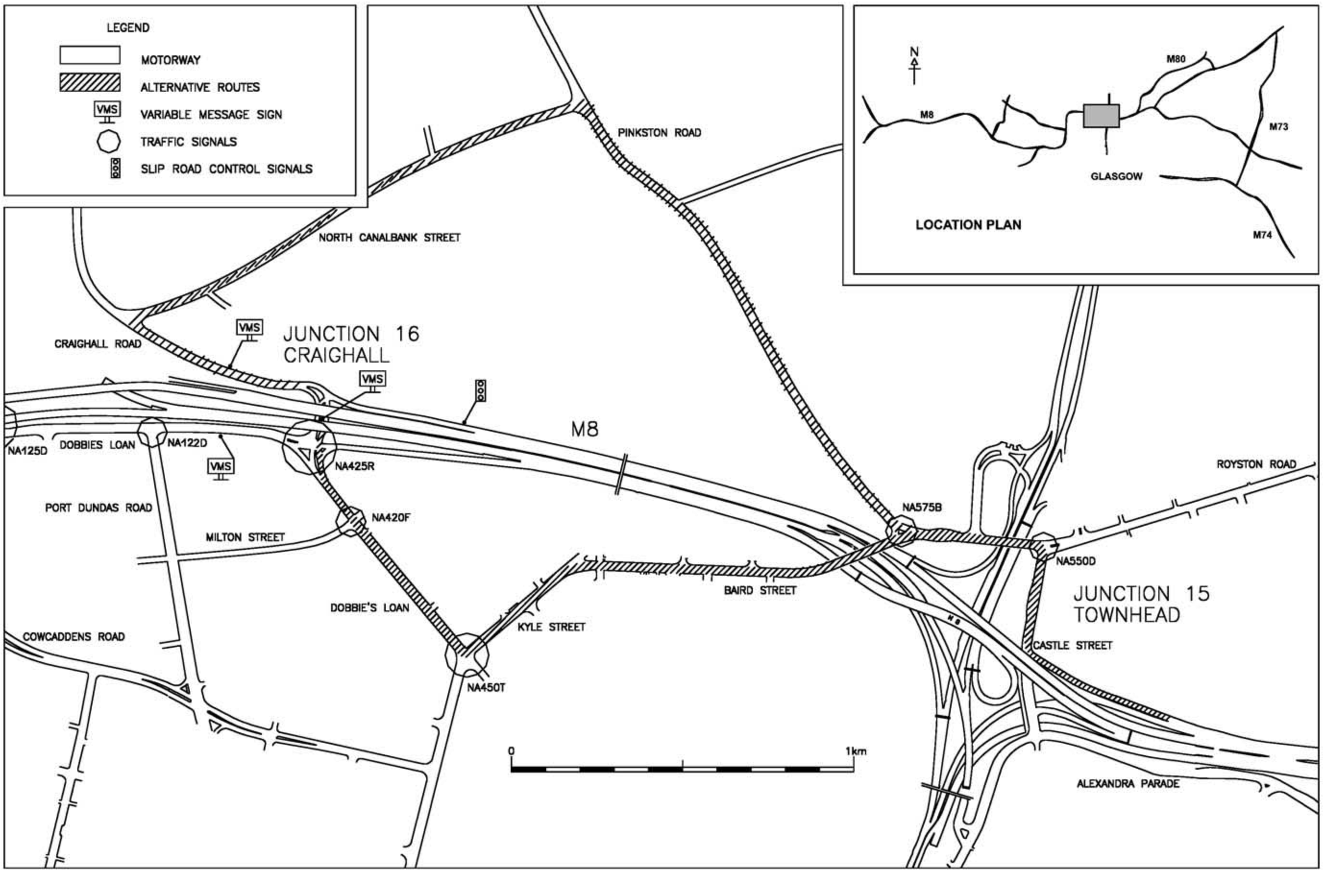


Fig. 8. The Glasgow site (McLean et al., 1997a).

Table 2
Effect of control on throughput and journey times (McLean et al., 1998)

	Time period: 16:00–17:00							
	Throughput				Journey times			
	veh/h	Percentage change			s	Percentage change		
	Base case	Case 1	Case 2	Case 3	Base case	Case 1	Case 2	Case 3
M8E motorway	36 721	+5	+6	+6	210	–5	[a]	[a]
Urban diversion routes	3087	+13	+20	+23	440	+4	[a]	+1
Total urban network	20 157	–3	–10	–6	1174	[a]	–11	–15
Total evaluation network (including the metered on-ramp)	60 324	+2	[a]	[a]	1567	–2	–10	–13

[a] Not statistically significant.

fully operational in February 1998. The evaluation was completed in March 1998. This approach allowed an understanding of the effect of individual control elements upon the network. The evaluation of the system considered a network wider than the one actually manipulated by the control system. This was done in order to enable a network-wide analysis of the effects of the system rather than an analysis constrained to the area directly affected by the operation of the system.

The main overall conclusions drawn from the multi-month field evaluation were (McLean et al., 1998):

- User acceptance was positive as indicated in their answers to distributed questionnaires.
- IN-TUC consistently increased the capacity of the motorway and the urban diversion routes.
- Throughput in the controlled corridor has significantly increased.

- The travel times have decreased in the case of the motorway and very slightly increased in the case of the urban network.
- Safety on the motorway is improved as the number of early merges of on-ramp vehicles is decreased along with the percentage of motorway lane changes upstream of the ramp.

Table 2 (McLean et al., 1998) summarises some of the performance indicators calculated throughout the evaluation period for the Glasgow corridor network. The 'M8E Motorway'-line refers to the eastbound direction of the Glasgow M8 motorway, the 'Urban Diversion Routes'-line refers to the directions of the roads of the alternative urban routes that lead to the main M8 eastbound exit, the 'Total Urban Network'-line refers to the whole considered urban network including the 'Urban Diversion Routes', while the 'Total Evaluation Network'-line refers to the sum of the mentioned network parts including also the metered on-ramp link. Furthermore, the 'veh/h'-column displays summations of traffic volume data from all available detectors in the corresponding network parts, while the 's'-column displays summations of journey times of the motorway and urban stretches of the corresponding network parts. Moreover, the evaluation considered several cases from which the following four main cases are presented in Table 2:

- *Base case*: Fixed urban traffic control only, without ramp metering nor VMS operation.
- *Case* 1: Fixed urban traffic control and ALINEA ramp metering.
- *Case* 2: ALINEA ramp metering and TUC signal control.
- *Case* 3: IN-TUC with ramp metering, signal control, and route guidance.

A case-by-case comparison in Table 2 reveals that:

- The introduction of ramp metering (case 1) increases (5%) the throughput of the motorway while decreasing (5%) the journey times therein. At the same time, having fixed urban traffic control at the urban diversion routes, the load (due to the drivers' diversion) and the journey times of the urban diversion routes increase significantly (13% and 4%, respectively).
- With the introduction of TUC strategy (case 2), the load of the urban diversion routes increases further (from 13% to 20%). In contrast, the journey times of the urban diversion routes decrease as compared to case 1 (from 4% to 0%) and return to the levels of the base case. From this, it may be concluded that the observed increased throughput and the amelioration of the journey times despite this increase are due to the efficient operation of the TUC strategy.
- The introduction of VMSs (case 3) leads to a further load of the urban diversion routes (from 20% to 23%), as in this case the VMSs may explicitly suggest the alternative routes to the drivers. However, the corresponding journey times present only a slight increase (from 0% to 1%), a fact that further supports the previous conclusion.

In conclusion, TUC specifically and IN-TUC as a whole demonstrated high efficiency in all field-investigated cases.

6. Conclusions and future work

The paper presents the traffic-responsive co-ordinated urban network control strategy TUC. TUC has been developed through the formulation of the urban traffic control problem as an optimal control problem based on a store-and-forward modelling approach. The employment of this modelling approach has two consequences for the derived strategy: (1) the control interval cannot be shorter than the cycle times of the considered signalised junctions, hence real-time decisions cannot be taken more frequently than at the maximum employed signal cycle; (2) the effect of offset for consecutive junctions cannot be directly considered. Additionally, the time-variance of traffic parameters like the turning rates and the saturation flows cannot be taken into account when deriving the multivariable regulators (7), (10), and (11). Finally, the centralised functional architecture of TUC and the multivariable regulator structure of the control law do not allow for immediate consideration of modifications and expansions of the controlled network.

Regarding the traffic parameters (turning rates, saturation flows), extensive simulation investigations (Diakaki, 1999) showed that, although the control matrices $\mathbf{L}$, $\mathbf{L_1}$ and $\mathbf{L_2}$ depend on them, they present a low sensitivity to their changes, hence the efficiency of the strategy remains practically unaffected. In case of modifications and expansions of the controlled network, the strategy has to be completely re-designed, i.e. the control matrices must be re-calculated. Nevertheless, the re-design is a straightforward task as it is performed using available generic software tools (Riccati equation) and exploits all the information that has come out from the initial design. The lack of flexibility of TUC regarding modifications and expansions of the controlled network is the price to be paid for its centralised functional architecture and the multivariable regulator structure of the control law that allow for the simultaneous consideration of all junctions with the application of a single and simple matrix equation independently of the network size.

Despite these minor shortcomings, the employed modelling and control approach has led to a highly efficient and extremely simple co-ordinated control strategy, applicable to large-scale networks as demonstrated in the presented simulation investigations, that carries also important features like robustness with respect to measurement inaccuracies, simplicity and transparency of the real-time code, and generality so that it may be transferred with minor modifications to networks with arbitrary topology and characteristics.

TUC has been implemented (Diakaki et al., 1999) and is currently operational in a part of Glasgow's (Scotland) urban network with excellent results. Furthermore, simulation investigations of TUC application in the city of Chania (Greece) have given extremely promising results (Dinopoulou, Diakaki, & Papageorgiou, 2000) leading to a field implementation that is currently under way.

Future research activities aim at comparing TUC with other urban traffic control strategies and at developing additional algorithms for bus-priority as well as for the real-time modification of the signal cycles and the offsets. Field application of TUC to extended network parts of three European cities is currently under way.

Acknowledgements

This work was partially supported by the European Commission under the Telematics Applications in Transport Project TABASCO (TR1054). The contents of the paper reflect the views of the authors, who are responsible for the accuracy of the data presented herein. The contents do not necessarily reflect the official policy of the European Commission.

References

Boillot, F., Blosseville, J. M., Lesort, J. B., Motyka, V., Papageorgiou, M., & Sellam, S. (1992). Optimal signal control of urban traffic networks. *Proceedings of the sixth IEE international conference on road traffic monitoring and control* (pp. 75–79). London, England.

D'Ans, G. C., & Gazis, D. C. (1976). Optimal control of oversaturated store-and-forward transportation networks. *Transportation Science*, *10*, 1–19.

Diakaki, C. (1999). *Integrated control of traffic flow in corridor networks*. Ph.D. thesis, Department of Production Engineering and Management, Technical University of Crete, Chania, Greece.

Diakaki, C., Papageorgiou, M., & McLean, T. (1997). Simulation studies of integrated corridor control in Glasgow. *Transportation Research C*, *5*, 211–224.

Diakaki, C., Papageorgiou, M., & McLean, T. (1999). Application and evaluation of the integrated traffic-responsive urban corridor control strategy IN-TUC in Glasgow. *Preprints CD-ROM of the 78th annual meeting of the transportation research board* (paper no. 990310). Washington, DC, USA.

Dinopoulou, V., Diakaki, C., & Papageorgiou, M. (2000). Simulation investigations of the traffic-responsive urban control strategy TUC. *Proceedings of the third annual IEEE conference on intelligent transportation systems* (pp. 458–463), Dearborn, MI, USA.

Dion, F., & Yagar, S. (1996). Real-time control of signalised networks—different approaches for different needs. *Proceedings of the eighth IEE international conference on road traffic monitoring and control* (pp. 56–60). Savoy Place, London, UK.

Elloumi, N., Haj-Salem, H., & Papageorgiou, M. (1994). METACOR: a macroscopic modelling tool for urban corridor. *TRISTAN* (*Triennal symposium on transportation analysis*) *II*, Capri, Italy.

Farges, J.-L., Henry, J.-J., & Tufal, J. (1983). The PRODYN real-time traffic algorithm. *Proceedings of the fourth IFAC symposium on transportation systems* (pp. 307–312). Baden-Baden, Germany.

Gartner, N. H. (1983). OPAC: a demand-responsive strategy for traffic signal control. *Transportation Research Record*, *906*, 75–84.

Gazis, D. C. (1964). Optimum control of a system of oversaturated intersections. *Operation Research*, *12*, 815–831.

Gazis, D. C., & Potts, R. B. (1963). The oversaturated intersection. *Proceedings of the second international symposium on traffic theory* (pp. 221–237). London, UK.

Hunt, P. B., Robertson, D. L., Bretherton, R. D., & Royle, M. C. (1982). The SCOOT on-line traffic signal optimisation technique. *Traffic Engineering and Control*, *23*, 190–199.

Lowrie, P. R. (1982). SCATS: The Sydney co-ordinated adaptive traffic system—principles, methodology, algorithms. *Proceedings of the IEE international conference on road traffic signalling* (pp. 67–70). London, UK.

Mauro, V., & Di Taranto, C. (1989). UTOPIA. *Proceedings of the sixth IFAC/IFIP/IFORS symposium on control, computers, communications on transportation* (pp. 245–252). Paris, France.

McLean, T., Brader, C., Diakaki, C., Papageorgiou, M., Hangleiter, S., Tsavahidis, M., Damas, C., & Maxwell, B. (1997a). *Sites, user needs and traffic model verification report*. Deliverable 8.1, Project TABASCO (TR1054), Transport Telematics Office, Brussels, Belgium.

McLean, T., Brader, C., Diakaki, C., Papageorgiou, M., Hangleiter, S., Tsavahidis, M., & Damas, C. (1997b). *Urban integrated traffic control implementation strategies*. Deliverable 8.2, Project TABASCO (TR1054), Transport Telematics Office, Brussels, Belgium.

McLean, T., Brader, C., Hangleiter, S., Tsavahidis, M., Damas, C., Maxwell, B., & Barber, P. (1998). *Urban integrated traffic control evaluation results*. Deliverable 8.3, Project TABASCO (TR1054), Transport Telematics Office, Brussels, Belgium.

Mirchandani, P., & Head, L. (1998). RHODES: a real-time traffic signal control system: Architecture, algorithms, and analysis. *TRISTAN* (*Triennal symposium on transportation analysis*) II, San Juan, Puerto Rico.

Papageorgiou, M. (1998). Automatic control methods in traffic and transportation. In M. Labbé, G. Laporte, K. Tanczos, & P. Toint (Eds.), *Operations research and decision aid methodologies in traffic and transportation management* (pp. 46–83). Berlin, Heidelberg: Springer.

Papageorgiou, M., Haj-Salem, H., & Blosseville, J.-M. (1991). ALINEA: a local feedback control law for on-ramp metering. *Transportation Research Record*, *1320*, 58–64.

Pavlis, Y., & Papageorgiou, M. (1999). Simple decentralized feedback strategies for route guidance in traffic networks. *Transportation Science*, *33*, 264–278.

Siljak, D. D. (1991). *Decentralized control of complex systems*. New York: Academic Press.

Singh, M. G., & Tamura, H. (1974). Modelling and hierarchical optimisation of oversaturated urban traffic networks. *International Journal of Control*, *20*, 269–280.