EVALUATION OF A DYNAMIC SIGNAL OPTIMISATION CONTROL MODEL USING TRAFFIC SIMULATION

Suphasawas NIGARNJANAGOOL
Intelligent Transport Systems Research Laboratory
The University of Queensland
Brisbane, Australia

Hussein DIA
Intelligent Transport Systems Research Laboratory
The University of Queensland
Brisbane, Australia

(Received December 15, 2004)

The objective of this paper is to demonstrate the feasibility of implementing a traffic signal optimisation model to improve real-time operations of traffic control systems. Advanced computer algorithms and traffic optimisation techniques can provide benefits over existing systems by reducing delays, improving travel times and reducing environmental emissions. The feasibility of the proposed approach is demonstrated by interfacing the traffic signal optimisation model to a microscopic traffic simulation tool, which enabled the evaluation of the benefits of the algorithm using computers in a controlled environment without disrupting traffic conditions. The main advantage of the proposed algorithm is its ability to detect dynamic changes in traffic flow conditions by using short-term historical demand data obtained from upstream vehicle loop detectors. The experimental results for under-saturated traffic conditions showed that the algorithm's performance was superior to optimal fixed time control. The results also confirmed that as traffic volumes reach saturated conditions, the performance of the algorithm decreased but remained better than what can be achieved by fixed time control systems.

Key Words: Traffic signal control and optimisation, Microscopic traffic simulation, Adaptive Traffic Management

1. INTRODUCTION

There is sufficient evidence in the literature to suggest that the application of advanced technologies including computers, electronics and communications can contribute to improving traffic conditions, enhancing environmental quality and increasing economic productivity. Collectively known as Intelligent Transport Systems, these technologies are rapidly being accepted by road and transport authorities around the world as a viable alternative to reliance on building more roads to reduce congestion. Advanced Traffic Management Systems (ATMS), in particular, have been shown in many experiments around the world to reduce travel times, improve travel time reliability for public transport systems, improve network speeds, reduce environmental emissions and congestion. The benefits of these systems have also been found to be a function of the accuracy and robustness of the underlying computer algorithms and optimisation techniques which provide various levels of intelligence to traffic signal control and operations.

Traffic signal control is generally defined as "power-operated traffic devices which alternatively direct traffic to stop and to proceed". More specifically, traffic signals are used to control the assignment of right-of-way at locations where conflicts exist or where passive devices, such as signs and markings, do not provide the necessary flexibility of control to properly move traffic in a safe and efficient manner. Traffic signal control is a key determinant for efficient operation of the urban street network and as such is considered an essential element of Intelligent Transport Systems.

Many cities around the world still implement fixed-time control traffic signal systems. These systems operate on a number of fixed or predetermined plans which are put in operation at different times of the day e.g., one set of plans would operate during morning peak, another during off peak and a third set of plans during evening peak hours. The disadvantage of these systems is that they do respond to changes in traffic demands and assume that traffic conditions during each of these periods will not change. If an accident happens between any two set of intersections, the traffic signal does not have the capacity to detect these changes or respond to such events.

Advances in computer technologies and communications systems have now allowed for the introduction of various levels of intelligence in these systems by enabling them to collect traffic data about flows, speeds and travel times to enable the system to respond to incidents, accidents, road works and other events that may reduce the capacity of the road system and contribute to congestion. Examples of such technologies include commercially available adaptive traffic management systems such as

SCATS and SCOOT in addition to other technologies being developed in research institutions such as Agent-based traffic management systems being developed in the ITS Research Laboratory at the University of Queensland[1] and the Traffic Responsive Urban Control (TRUC) system being developed at the Dynamic Systems and Simulation Laboratory, Technical University of Crete in Greece[2]. The objective of this paper is to contribute to such research and development efforts by demonstrating the application of a signal optimisation model, which is generally used to design pre-timed traffic signal timings, to real-time network-wide optimisation of traffic conditions. Interestingly, the origins of this signal optimisation model are found in the work by Webster in 1958[3]. However, the model was initially developed for operation in under-saturated conditions[4]. A number of researchers introduced several changes to the model over the years and a variant of the model (described below) is currently implemented in the well-known aaSIDRA[4,5] software.

To demonstrate the applicability and potential of the technique, this study evaluated the algorithm using a microscopic traffic simulation model, AIMSUN[6], on a real world network in Brisbane comprising 28 signalised intersections. A brief description of the AIMSUN model is provided in Section 4.

2. REVIEW OF SIGNAL OPTIMISATION MODELS

Over the years, traffic engineers have used several methods for designing pre-timed isolated signals. For example, Homburger and Kell's method utilised traffic volumes as the basis for allocating times to approaches with the constraint of keeping off-peak cycles as short as possible (40 to 60 seconds)[7]. The highway capacity method, on the other hand, determines the traffic signal cycle length based on the capacity of lane groups[8].

The well known method which influenced the Australian and U.K. signal design practice is Webster's method[3], which he introduced to obtain an optimum cycle time that produces minimum delays to vehicles. He proposed the use of a "lost time" parameter (which represents the time lost before vehicles start to move) and the saturation flow (which is the maximum rate of discharge). The Webster method has since become a traditional technique to design signal timings for isolated intersections both in Australia and overseas[4,5,9].

In 1984, Akcelik[4] introduced several changes to the original formulation. The main modification included changing the core concept of 'phase-related' methods to 'movement-related' techniques. Consequently, an important aspect of this change was the use of 'movement lost time' instead of 'phase lost time' which led to a definition of the intersection's lost time as 'the sum of critical movement lost times' rather than 'the sum of phase lost times'. This new approach also facilitated a clearer understanding of the relationships between movement and signal phasing characteristics.

3. MOVEMENT-BASED SIGNAL TIMING OPTIMISATION MODEL

This study applies the movement-related method proposed by Akcelik to determine traffic signal cycles. A brief explanation of the approach, as implemented in this study, is presented next. Unfortunately, a full discussion of the technique is outside the scope of this paper but the reader is referred to Akcelik[4] for a detailed explanation of the method.

3.1 Movement characteristics

The basic movement characteristics are illustrated in Figure 1. The main movement parameters are saturation flow, effective green time and lost time. The model assumes that when the signal changes to green, the flow across the stop line increases rapidly to saturation flow which remains constant until either the queue is exhausted or the green period ends. As indicated by the dotted line in Figure 1, the model replaces the actual departure flow curve by a rectangle of equal area whose height is saturation flow (s) and width is the effective green time (g).

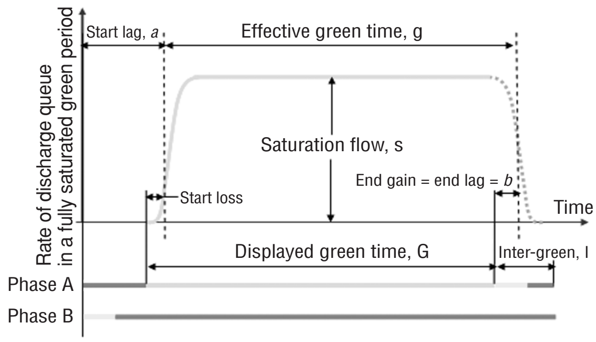


Fig. 1 Basic movement characteristics[3-5]

The start and end times of the effective green period for a movement are best defined with reference to phase change times. Start lag (a) is defined as the sum of the movement inter-green time (I) and start loss, and end lag (b) is defined simply as the end gain. The difference between the start and end lag time is defined as movement lost time (l). It should be noted that the movement inter-green time is the inter-green time of the starting phase of the movement.

$l = a - b$ .... (1)

From Figure 1, the relationship between the displayed green time (G) and the effective green time (g) is

$g + l = G + I$ .... (2)

The sum of all phase inter-green and green times is the cycle time (c):

$c = \Sigma (I + G)$ .... (3)

From Equation 2 and Equation 3, the similar relationship holds for movement parameters is:

$c = \Sigma (g + l)$ .... (4)

where the summation is for critical movements.

The movements which determine the capacity and timing requirements of the intersection are called critical movements. Sufficient time must be allocated to each critical movement to meet its capacity requirements in order to give all movements sufficient capacity. The technique to identify the critical movements is explained in Section 3.2.

The time allocated to a movement is the sum of effective green time (g) and lost time (l) and is given by:

$t = g + l = I + G$ .... (5)

The required movement times can be calculated from:

$t = 100\, u + l$ .... (6)

100 is the first estimate of cycle time and u is the required green time ratio which is calculated to achieve maximum acceptable (practical) degree of saturation (x_p), and is given by:

$u = y / x_p$ .... (7)

Movement flow ratio (y) is the ratio of arrival flow (q) to saturation flow (s) is given by:

$y = q / s$ .... (8)

It must be noted that the movement time calculated from Equation 6 must satisfy the sum of fixed minimum effective green time and lost time as:

$t = g_m + l$ .... (9)

The minimum displayed green time (G_m), therefore, can be calculated from the relationship:

$g_m + l = G_m + I$ .... (10)

3.2 Critical movement identification

The identification method is based on the comparison of the required movement time (t) values. If all movements were non-overlap movements, there would be one critical movement per phase. This would be the movement which requires the longest movement time in the phase. For the overlap movements, their movement time includes the green and inter-green times of all phases during which it has right of way. The overlap movement is critical if its t value is longer.

This method requires a phase-movement matrix as shown in Table 1, and a critical movement search diagram as illustrated in Figure 2. In the diagram, the nodes correspond to phase change events, and the links to movements.

Table 1 Phase-movement matrix

Movement	Starting Phase	Terminating Phase
1	A	C
2	A	B
3	B	C
4	C	A

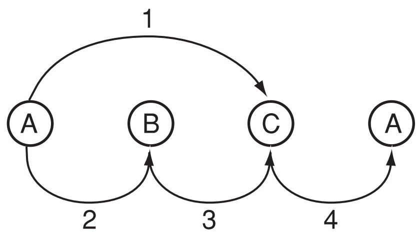


Fig. 2 Critical movement search diagram

Then, intersection lost time (L), intersection green time ratio (U) and intersection flow rate (Y) which are the summation of respective critical movement parameters are calculated.

3.3 Cycle and phase timing determination

The optimum cycle time, which minimises a performance measure that is a function of total delays and number of stops for all critical movements at an isolated intersection, is calculated from the following formula:

$c_o = (1.4{+}k)\, L + 6 / (1 - Y)$ .... (11)

where k = K/100 is the stop penalty parameter. This study selected a value of k which is commonly used to minimise cost (k = 0.2). The practical cycle time which ensures that the degrees of saturation of all movements are below specified maximum acceptable degree of saturation, $x < x_p$, is calculated from the following formula:

$c_p = L / (1 - U)$ .... (12)

This paper selected a default value of $x_p = 0.9$.

For a given cycle time (c) and total available green time ($c - L$), the available green time can be distributed to critical movements according to the formula:

$g = u\,(c - L)\,/\,U$ .... (13)

For non-overlap movements, where the overlap movement is critical, the green time for non-overlap movements can be calculated by treating the critical movement time as a sub-cycle time, $c^* = g_c + l_c$, where g_c and l_c are the critical movement green and lost times. Then available green time is ($c^* - L^*$), where L^* is the sum of non-overlap movement lost times. The green time distributed to non-overlap movement is as follows:

$g = u\,(c^* - L^*)\,/\,U^*$ .... (14)

By modifying Equation 5, the displayed green time for a phase can be calculated from:

$G = (g + l) - I$ .... (15)

where ($g + l$) is the time allocated to a movement which receives right of way during the phase only, and I is the inter-green time of that phase.

Having introduced the reader to the theoretical background of the algorithm, the next section of this paper presents the application of this algorithm within microscopic traffic simulation to evaluate its performance under a variety of traffic conditions and scenarios.

4. APPLICATION DEVELOPMENT

A number of commercially available traffic simulation models provide facilities to interface the simulator to an external algorithm (such as the one being tested in this study) to enable a full evaluation of its performance under controlled conditions. For example, PARAMICS[10] provides a Programmer module and AIMSUN[6] provides GETRAM Extensions for developing a plug-in module in C++ code. This study employed the AIMSUN for developing, testing and evaluating the application. This section presents the model development, data requirements and traffic modelling tasks needed to set up the simulator and interface it to the signal optimisation model.

4.1 Model development and data requirements

This study used the Brisbane Western Corridor Traffic Simulation Model as a test-bed for evaluating the performance of the signal optimisation model. The modelled traffic network, shown in Figure 3, was developed based on the year 2002 traffic conditions. The study area was about 3.1km (East-West) and 2.1km (North-South) or approximately 6.5 square kilometres. The lengths of the major routes in the study area were 3.3 kilometres for Coronation Drive and 3.2 kilometres for Milton Road.

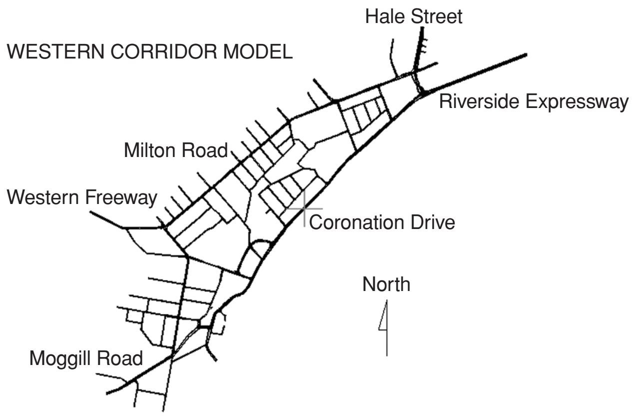


Fig. 3 Schematic of the Brisbane Western Corridor Model

Microscopic simulation is characterised by a high level of modelling detail. The accuracy of the model will depend on the availability and quality of the input data. The following data were collected and used for model development in this study:

Network Layout – Digitised maps (DXF format) and raster images showing locations of both signalised and unsignalised intersections, possible turning movements for each intersection, recommended turning speeds for each movement, visibility distances at junctions, presence of Stop or Yield signs and detector positioning were required. For each section, the road centrelines, number of lanes in each section of road, width of each lane, reserved lanes (where entry is allowed only for certain vehicle types), restrictions on lane changes (solid line markings), maximum speed for each section (or each lane if necessary), capacity (vehicles per hour for use in cost functions for the calculation of shortest paths), visibility distances, length of each section, slope and lane changing distances were also required. A schematic of the network is shown in Figure 3.

Traffic Demand Data – In AIMSUN, traffic demand data can be defined either by the combination of traffic flows at input sections and turning proportions at intersections or by Origin-Destination (O-D) matrices. For this study, the traffic demand on the network was represented by O-D matrices related to centroids. The number of trips between each O-D pair were required for each vehicle type in time-sliced intervals (e.g., 15 minutes). The O-D matrices gave the number of trips from each origin centroid to each destination centroid, for each time slice (e.g. 15 minutes) and for each vehicle type.

As was discussed previously, basic movement characteristics including inter-green (I), minimum displayed green time (G_m), saturation flow (s) and practical degree of saturation (x_p) are all required as inputs to the model. Default values were used for these inputs[4]. A value of 0.9 was selected as practical degree of saturation for all movements of all intersections.

The application also required the input of traffic volumes for the network. The real-time traffic volumes were obtained by vehicle loop detectors located at the stop-line of every signalised intersection. However, in order to optimise signal timings for the next cycle, the predicted traffic volume is also required. The application used the average traffic volumes from three previous cycles to capture the change in traffic condition.

The Origin-Destination (O-D) matrix used in this study was for the morning peak starting from 6:45 A.M. to 8:45 A.M.

4.2 Simulator settings

The use of traffic simulation software requires that several simulation parameters are specified. These include important parameters related to car following and lane changing algorithms and other global parameters that control the conduct of the experiments. Some of the important parameters used in the study included reaction time, simulation step per second and reaction time at stop. All of these parameters were assigned a value of 1 second. For route choice, a fixed distance was selected as the default for all experiments. This meant that route choice was not an option and that all vehicles between origins and destinations followed the same route. The simulation was run for five replications to ensure statistical reliability. The simulation time was two hours with a 30-minute warm-up period to populate the network with vehicles and allow traffic conditions to stabilise before collection of statistics on performance.

5. EVALUATION OF THE DYNAMIC SIGNAL OPTIMISATION ALGORITHM USING TRAFFIC SIMULATION

Having developed, calibrated and validated the traffic simulation model, the next task was to interface the signal optimisation model to the traffic simulator to evaluate its performance. The signal optimisation algorithm was evaluated in two different scenarios. First, the algorithm was used to control the isolated intersections assuming that the effects of upstream and downstream traffic from adjacent intersections are negligible. Second, the algorithm was applied to control the 17 signalised intersections on Coronation Drive (which is the main route connecting the western suburbs to the CBD as was shown in Figure 3).

5.1 Scenario 1: Isolated junction without effects from adjacent intersections

The Toowong junction in Figure 4 was selected since it had a relatively large traffic volume and one of its approaches is a short link (shown as approach number 1 in Figure 4). In the morning peak, there is large traffic volume moving from Approach 1 and Approach 3 through this intersection to city.

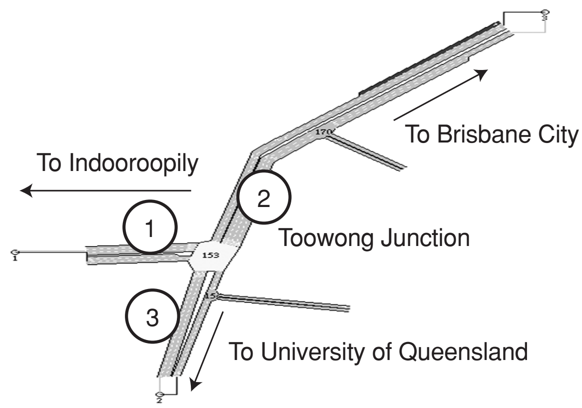


Fig. 4 Toowong junction in Brisbane

The performance of the dynamic signal optimisation model was compared to an optimum fixed cycle time setting obtained by aaSIDRA[10]. The comparative results are shown in Figures 5 and 6 where the dotted and solid lines represent dynamic and optimum fixed time control, respectively. The simulations were conducted for two hours simulation time with $v/c = 1.0$ (i.e the volume was set equal to capacity) and the majority of traffic was moving from approach 1 to the city (left turn traffic from 1 to 2).

The plots shown in Figure 5 clearly show how the dynamic cycle control algorithm increased the throughput across the intersection, especially during medium to heavy conditions, when compared to fixed time control. This demonstrates how simple improvements to the underlying core computer algorithms which control the operation of traffic signal systems can produce substantial

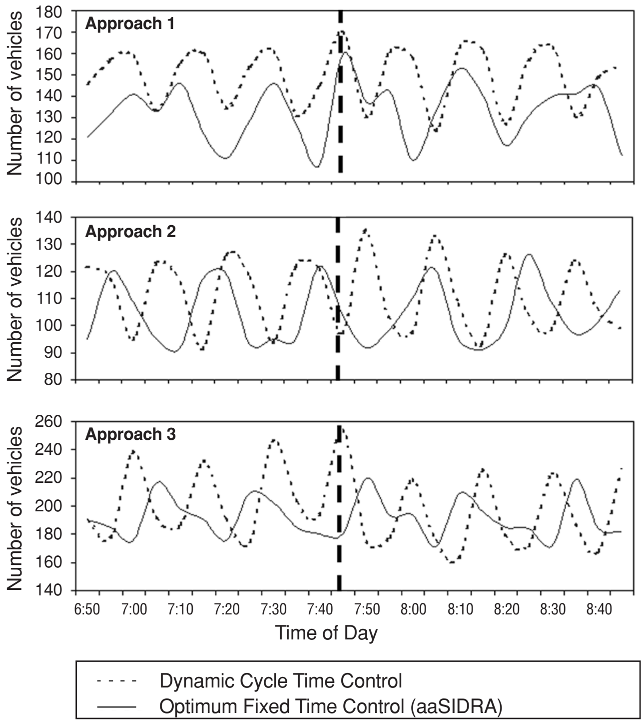


Fig. 5 Comparative results of dynamic and optimum fixed cycle time by approach

benefits without the need to spend large amounts of funds to build or construct new roads to improve capacity.

Figures 5 and 6 also show that dynamic cycle time control was superior to optimum fixed time control as traffic volumes started to increase (e.g., around 7:45 A.M.). In Figure 6(a), the dynamic cycle time control reduced delays by approximately 20 percent and produced higher average speeds as shown in Figure 6(b). In the beginning of simulation, there were few vehicles entering the junction. Under fixed time control most of the vehicles were able to pass through the junction without queuing. Since the dynamic cycle time control attempts to optimise cycle time based on predicted traffic volumes, the cycle times during the beginning periods resulted in a larger number of stops as shown in Figure 6(c).

The dynamic cycle time control initially attempted to provide sufficient time for the major movements to pass through the junction. This clearly resulted in queues forming at the other approaches with less demand, where some of the vehicles on the minor approaches waited for more than a cycle to clear the intersection. The number of stops under dynamic cycle time control was reduced to less than two.

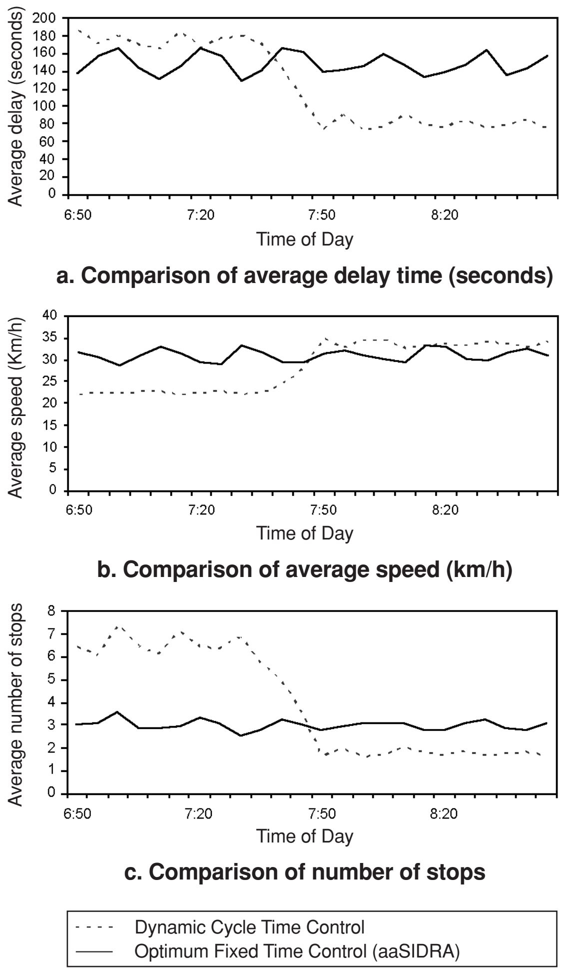


Fig. 6 Comparative results of dynamic and optimum fixed cycle time control

In fact, it was observed that the controller was reducing the queue length by a small amount in every cycle by adjusting the cycle time and phase proportions to suit the new traffic demands for the next cycle. In other words, while fixed time control was operating the intersection using the same cycle and phase time proportions, the dynamic cycle time control was adjusting the time and proportions to suit the current demand for all movements as shown in Figures 7 and 8.

Figure 9 presents sensitivity analysis results for the two control strategies when the degree of saturation was varied from light traffic (v/c= 0.4) to over saturated traffic conditions (v/c > 1.0). These figures show that dynamic cycle control which uses real-time traffic volumes to determine signal timings produced less delay in light traffic conditions (v/c = 0.4 to 0.7). However, both control strategies performed in a similar manner when traffic volumes approached capacity.

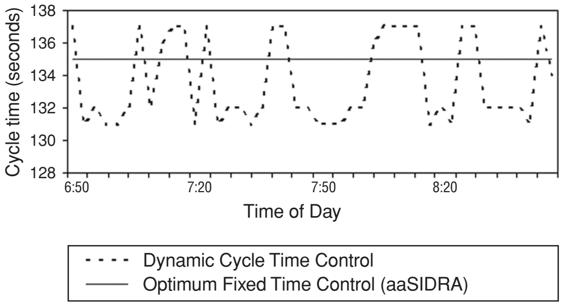


Fig. 7 Variability in cycle times between 6:45 A.M. to 8:45 A.M.

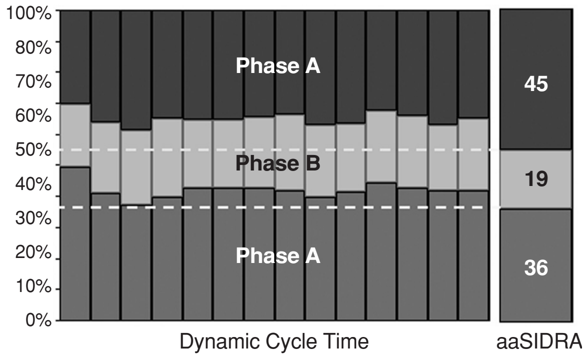
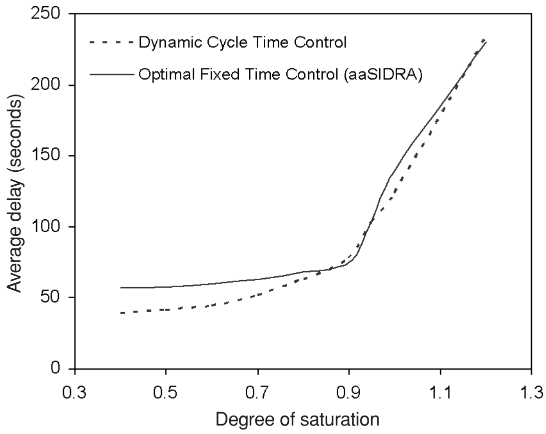
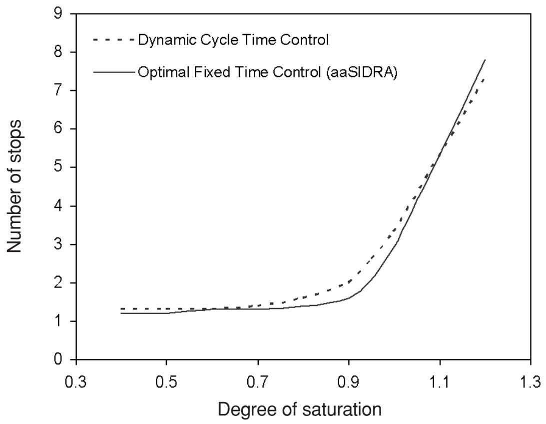


Fig. 8 Phase time proportions



a. Average delay for variable degrees of saturation



b. Average number of stops for varied degrees of saturation

Fig. 9 Results of sensitivity analysis

5.2 Scenario 2: Commuting corridor with effects from adjacent intersections

In this scenario, the dynamic signal control was used to operate 17 out of the 28 signalised intersections in the Brisbane Western Corridor traffic network, as illustrated in Figure 10. More than 5,000 vehicles per hour use this route during the morning peak. The simulation was run for two hours with morning peak O-D matrix. In order to allow the same traffic volume for both control strategies, the route choice model option in AIMSUN was set as a fixed distance. This meant that route choice was disabled to allow the effects of the control strategy to be evaluated.

In this scenario, the algorithm was evaluated at intersections where there are effects from upstream and downstream intersections, which is more realistic and common than Scenario 1. This meant that the queue formations from downstream intersections could create spillback or lane blocking conditions, which may result in unused green time and unnecessary delays. The dynamic cycle time control addressed this by considering measurements from loop detectors which helped detect whether

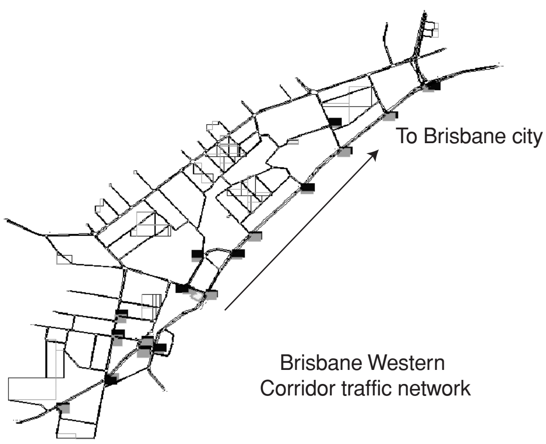


Fig. 10 Commuting corridor in Brisbane

traffic was passing through the intersection. The algorithm then assigned the minimum green time for that particular movement until conditions changed. It should be noted here that phase termination was not considered as an option is this study.

The comparative results of both control strategies are shown in Figure 11 where the dotted and solid lines represent dynamic and fixed time control, respectively. Five replications were run for statistical reliability and the periodic statistical results were recorded every 15 minutes. The simulation time comprised three periods (light, medium and heavy traffic conditions).

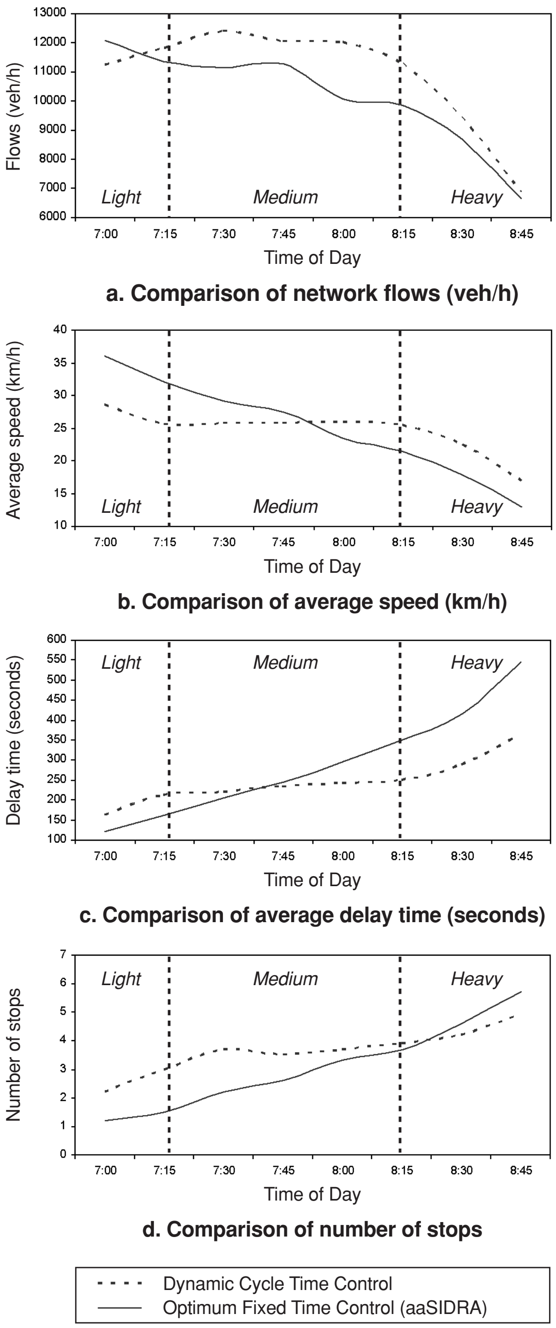


Fig. 11 Comparative results of dynamic and fixed time control

During light traffic conditions (before 7:15 A.M.), the dynamic cycle time control was aiming to provide sufficient time for all movements which resulted in delays on the approaches with less demand. The dynamic cycle control allocated green times to suit current traffic demands while fixed time control provided a constant green time. During medium traffic conditions (after 7:15 A.M.), the dynamic cycle time control performed better than fixed time plans. The results presented in Figure 11 also indicate that, on average, speeds, delays and number of stops for the dynamic control scenario were superior to fixed time control (e.g. delays were reduced by more than 10 percent). It should be noted here that the improvements were not better than the first scenario due to the effects from upstream and downstream intersections.

Under heavy traffic conditions (after 8:15 A.M.), the network was full of vehicles and the volume exceeded capacity. The performance of the dynamic control strategy decreased but remained superior to fixed time control. It should be noted here that this is not a unique characteristic to this optimisation algorithm. Most adaptive traffic management systems (e.g., SCATS and SCOOT) also demonstrate similar patterns where the control strategy's benefits during over saturated conditions are essentially limited and are similar to what can be achieved by fixed time systems.

It is also noted here that the main objective of this study was to demonstrate the benefits of applying the dynamic cycle time optimisation algorithm to individual intersections. Although not considered in this study, it is expected that implementation of a coordination strategy using offset optimisation can further improve the results for the dynamic cycle time algorithm. Offset optimisation techniques will be considered in future studies and extensions of this work.

6. CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

The dynamic cycle time control presented in this paper is based on existing well established fixed cycle time optimisation techniques[6]. The contribution of this work was to evaluate the performance of the algorithm and its potential for optimising network-wide traffic conditions based on dynamic demand data provided by loop detectors. The dynamic control strategy was evaluated using the AIMSUN traffic simulator. The optimisation logic of the control strategy was based on the calculation of optimal signal timing and green time allocation for each movement at the end of every cycle by using predicted demands calculated based on average volumes from previous cycles. The application was evaluated in two different scenarios. In the first scenario, an isolated intersection which operated without influences from downstream intersections was explored. In the second scenario, the algorithm was tested on 17 signalised intersections along the Corridor. The results for both scenarios showed that the performance of the dynamic control system was superior to fixed time control and provided better throughput across the intersections. The performance of the dynamic control strategy started to decrease under very heavy traffic conditions but remained superior to fixed time control. This finding is consistent with results for most adaptive traffic management systems which show that their performance under saturated conditions is limited and similar to what can be achieved by fixed time control strategies. Further research is being conducted at the University of Queensland's ITS Research Laboratory to improve the performance of dynamic traffic signal control strategies using intelligent agents theories[1].

REFERNECES

1. Nigarnjanagool, S. Development and Evaluation of an Agent-Based Adaptive Traffic Signal Control System. Proceedings of 25th Conference of Australian Institutes of Transport Research (CAITR), The University of South Australia, Adelaide. (2003).
2. Diakaki, C., Dinopoulou, V., Aboudolas, K., Papageorgiou, M., Ben-Shabat, E., Seider, E. and Leibov, A. Extensions and New Applications of the Traffic Signal Control Strategy TUC. Proceedings of the 82nd Transportation Research Board Meeting, Washington DC. (2003).
3. Webster, F. V. & Cobee, B. M. Traffic signals. Road research technical paper 56. Her Majesty's Stationary office, London. (1962).
4. Akcelik, R. Traffic signal: capacity and timing analysis. Australian Road Research Report No.123. Australia. (1984).
5. aaTraffic. Introducing aaSIDRA 2.1. Web: http://www.aatraffic.com/downloads/Introducing_aaSIDRA.pdf. (2004).
6. Transport Simulation Systems. GETRAM/AIMSUN. Web: www.aimsun.com. (2004).
7. Homburger, W. S. Fundamentals of traffic engineering. University of California, Berkeley. (1996).
8. Garber, N. J. & Hoel, L. A. Traffic and highway engineering. Books/Cole Thomson learning, USA. (2001).
9. Kristy, C. J. & Lall, B. L. Transportation engineering: an introduction. Prentice Hall, New Jersey. (2003).
10. Quanstone. Paramics. Web: www.paramics-online.com. (2004).