A Non-homogenous Time Mixed Integer LP Formulation for Traffic Signal Control

Iain Guilliard
iguilliard@nicta.com.au

2254 words + 7 figures + 5 tables
June 26, 2015

INTRODUCTION

This report describes a new model for optimised traffic signal planning using a MILP formulation. Previously, Lin and Wang (1) describe a MILP formulation for optimised traffic signal planning based on the Cell Transmission Model (2). However a CTM based model is limited in scalability by the requirement that each road way in the network must be partitioned up into segments that take exactly $\Delta t = 1$ to traverse at the free flow speed. We get around this limitation by using a queue based model that supports non-homogeneous time steps, and we will show how we can exploit this property to scale a network without significant increase in the number of variables or loss of quality.

THE QUEUE TRANSMISSION MODEL

The Queue Transmission Model (QTM) consists of a network $\mathscr{N} = (\mathscr{Q}, \mathscr{L})$, where $\mathscr{Q}$ is a set of queues and $\mathscr{L}$ is a set of lights. A queue is defined by the tuple $(\mathrm{Q}, \mathrm{Q}^{\mathrm{in}}, \mathrm{Q}^{\mathrm{out}}, \mathrm{Q}^{\mathrm{delay}}, \mathbf{F}, \mathbf{Pr}, \mathscr{Q}^{\mathscr{P}})$, where $\mathbf{F}$ and $\mathbf{Pr}$ are vectors of maximum flow rates and turn probabilities into connected queues, $\mathscr{Q}^{\mathscr{P}}$ is the set of phases controlling the queue, and Q,Q^{in},$\mathrm{Q}^{\mathrm{out}}$, and $\mathrm{Q}^{\mathrm{delay}}$ are defined in table 1. Vehicles travel down the length of the queue at the free flow speed and enter into a vertically stacked queue. A light is defined by the tuple $(\Psi^{\min}, \Psi^{\max}, \mathscr{P})$, where $\mathscr{P}$ is the set of phases of the light and each phase is defined by the pair $(\Phi^{\min}, \Phi^{\max})$, and where $\Psi^{\min}$,$\Psi^{\max}$,$\Phi^{\min}$, and $\Phi^{\max}$ are defined in table 1. Examples of QTM networks are shown in figs. 3 to 5.

LP FORMULATION

We have defined a LP formulation for the QTM supporting non-homogenous Δt and a single binary variable per signal phase. The implementation also supports configurable phase and cycle constraints. The model constants are listed in table 1 and the variables in table 2

Network Constraints

For a set of time intervals over the period $[0, \mathrm{T}]$, where the duration of each interval n is Δt_n, we define a set of variables for each interval and each queue in the network in table 2 and the associated constants in table 1.

First, all the variables in table 2 are constrained to be ≥ 0. Next we constrain the external flows into and out of q_j during interval n,

$$in_{i,n} \leq \mathrm{Q}_i^{\mathrm{in}} \tag{C1}$$

$$out_{i,n} \leq \mathrm{Q}_i^{\mathrm{out}} \tag{C2}$$

and the internal flow from q_j to q_i,

$$f_{i\to j,n} \leq \mathrm{Pr}_{i\to j} \sum_{k=1}^{|\mathscr{Q}|} f_{i\to k,n} \tag{C3}$$

where we find the proportion of flow out of queue i turning into queue j by weighting the total flow by the turn probability $\mathrm{Pr}_{i\to j}$, noting that $\sum_k \mathrm{Pr}_{i\to k} = 1$. The flow out from queue i is controlled by the set of phases $\mathscr{Q}_i^{\mathscr{P}}$, so we modulate $f_{i\to j,n}$ by applying the constraint,

$$f_{i\to j,n} \leq \mathrm{F}_{i\to j} \sum_{p_{\ell,k,n} \in \mathscr{Q}_i^{\mathscr{P}}} p_{\ell,k,n} \tag{C4}$$

Constant	Desciption
$\lvert\mathscr{Q}\rvert$	number of queues
$\lvert\mathscr{L}\rvert$	number of lights
$\lvert\mathscr{P}_\ell\rvert$	number of phases of light ℓ
N	number of intervals
Δt_n	time duration of interval n
t_n	elapsed time at interval n
T	maximum elapsed time over all intervals (t^N)
Q_i	maximum capacity of queue i
$\mathrm{Q}_i^{\mathrm{in}}$	maximum inflow to queue i from outside the network
$\mathrm{Q}_i^{\mathrm{out}}$	maximum outflow from queue i to outside the network
$\mathrm{Q}_i^{\mathrm{delay}}$	propagation delay along queue i
$\mathrm{F}_{i\to j}$	maximum flow from queue i into queue j
$\mathrm{Pr}_{i\to j}$	proportion of total flow out from queue i into queue j (turn probability)
$\Phi_{\ell,k}^{\max}$	maximum allowed duration of phase k of light ℓ
$\Phi_{\ell,k}^{\min}$	minimum allowed duration of phase k of light ℓ
$\Psi_\ell^{\max}$	maximum allowed cycle time of light ℓ
$\Psi_\ell^{\min}$	minimum allowed cycle time of light ℓ

TABLE 1 Constants

Where $p_{ell,k,n}$ is 1 if the phase is active during interval n, and 0 otherwise. We can then sum the total flows in and out of each queue,

$$q_{i,n}^{\mathrm{in}} = in_{i,n}\Delta t_n + \sum_{j=1}^{|\mathscr{Q}|} f_{i\to j,n}\Delta t_n \tag{C5}$$

$$q_{i,n}^{\mathrm{out}} = out_{i,n}\Delta t_n + \sum_{j=1}^{|\mathscr{Q}|} f_{j\to i,n}\Delta t_n \tag{C6}$$

$$q_{i,n}^{\mathrm{out}} \leq q_{i,n} \tag{C7}$$

with the constraint C7 that the total flow out of queue i during interval n cannot exceed the volume of the queue at the start of that interval. Now we can perform the update step for queue i over interval n,

$$q_{i,n} = q_{i,n-1} - q_{i,n-1}^{\mathrm{out}} + (1-\alpha_{i,n})q_{i,m}^{\mathrm{in}} + \alpha_{i,n}q_{i,m+1}^{\mathrm{in}} \tag{C8}$$

$$q_{i,n} \leq \mathrm{Q}_i \tag{C9}$$

$$(1-\alpha_{i,n})q_{i,m}^{\mathrm{in}} + \alpha_{i,n}q_{i,m+1}^{\mathrm{in}} + \sum_{k=m+2}^{i} q_{n,n}^{\mathrm{in}} \leq \mathrm{Q}_i - q_{i,n-1} \tag{C10}$$

Variable	Type	Range	Description
$q_{i,n}$	continuous	$[0, \mathrm{Q}_i]$	traffic volume of queue i during interval n
$q_{i,n}^{\text{out}}$	continuous	$[0, \infty)$	outflow from queue i during interval n
$q_{i,n}^{\text{in}}$	continuous	$[0, \infty)$	inflow to queue i during interval n
$in_{i,n}$	continuous	$[0, \mathrm{Q}_i^{\text{in}}]$	inflow to the network via queue i during interval n
$out_{i,n}$	continuous	$[0, \mathrm{Q}_i^{\text{out}}]$	outflow from the network via queue i during interval n
$f_{i \to j,n}$	continuous	$[0, \mathrm{F}_{i \to j}]$	flow from queue i into queue j during interval n
$p_{\ell,k,n}$	binary	$\{0, 1\}$	signal phase k of light ℓ during interval n(1=green)
$d_{\ell,k,n}$	continuous	$[0, \Phi_{\ell,k}^{\max}]$	duration of phase k of light ℓ during interval n

TABLE 2 Variables

where m is the interval containing $t_n - \mathrm{Q}_i^{\text{delay}}$ such that,

$$t_m \leq t_n - \mathrm{Q}_i^{\text{delay}} < t_{m+1} \tag{1}$$

Since the model is piecewise linear, we linearly interpolate $q_{i,n}^{\text{in}}$ across the interval m to find the inflow to queue i at $t_n - \mathrm{Q}_i^{\text{delay}}$, and $\alpha_{i,n}$ is calculated in a pre-computation step for all i and all n,

$$\alpha_{i,n} = \frac{t_n - \mathrm{Q}_i^{\text{delay}} - t_m}{\Delta t_m} \tag{2}$$

Note that if $\mathrm{Q}_i^{\text{delay}}$ is a homogeneous number of time intervals, $n - m$, then $t_n - \mathrm{Q}_i^{\text{delay}} - t_n m = 0$ and constraint C8 reduces to

$$q_{i,n} = q_{i,n-1} - q_{i,n-1}^{\text{out}} + q_{i,m}^{\text{in}} \tag{3}$$

MILP FORMULATION

If we take $p_i^n \in \Pi(0, \mathrm{T})$, where Π is a fixed signal plan over all the intervals from 0 to T^{MAX}, then the constraints C1 to C10 form a dynamic, piecewise linear model of flow in the network over time as a function of Π. Alternatively we can define p_j^n as a binary variable and solve to find both a network flow and an optimal signal plan for a given objective function.

Phase constraints

First we map each p_i^n to a signal phase k of a light ℓ as $p_{\ell,k,n}$ (Note that there could be more than one queue mapped to each $p_{\ell,k,}$, or their could be none). Then we define a set of constraints for the signal phases. For each traffic light ℓ, we constrain the phases of ℓ such that exactly one phase is active in each interval n, and so that they activate sequentially,

$$\sum_{k=1}^{|\mathscr{P}_\ell|} p_{\ell,k,n} = 1 \tag{C11}$$

$$p_{\ell,k,n} + p_{\ell,k+1,n} \leq 1 \tag{C12}$$

$$p_{\ell,k,n-1} \leq p_{\ell,k,n} + p_{\ell,k+1,n} \tag{C13}$$

where $k+1 = 1$ if $k = P_l$. The constraints C12 and C13 ensure that if $p_{\ell,k,}$ was active during interval $n-1$ and has become inactive in interval n, then $p_{l,k+1}$ becomes active in interval n.

Next we enforce the minimum and maximum phase durations, $\Phi^{\min}_{\ell,k}$ and $\Phi^{\max}_{\ell,k}$ for each $p_{\ell,k,}$, by defining a duration variable $d_{\ell,k,}$ for each phase. When $p_{\ell,k,}$ is active, $d_{\ell,k,}$ holds the elapsed time since the start of phase k, and when phase k is inactive $d_{\ell,k,n}$ is constant and holds the duration of the last phase until the next activation,

$$d_{\ell,k,n} = \begin{cases} d_{\ell,k,n-1} + \Delta t_{n-1} & p_{\ell,k,n-1} = 1, p_{\ell,k,n} = 1 \\ d_{\ell,k,n-1} & p_{\ell,k,n} = 0 \\ 0 & p_{\ell,k,n-1} = 0, p_{\ell,k,n} = 1 \end{cases} \tag{4}$$

We achieve this by applying a set of linear envelope constraints, using the "big M" trick to activate each section of the envelope depending on the state of $p_{\ell,k,}$, where "big M" can be limited to $\Phi^{\max}_{\ell,k}$.

$$d_{\ell,k,n} \leq d_{\ell,k,n-1} + \Delta t_{n-1} p_{\ell,k,n-1} + \Phi^{\max}_{\ell,k}(1 - p_{\ell,k,n-1}) \tag{C14}$$
$$d_{\ell,k,n} \geq d_{\ell,k,n-1} + \Delta t_{n-1} p_{\ell,k,n-1} - \Phi^{\max}_{\ell,k}(1 - p_{\ell,k,n-1}) \tag{C15}$$
$$d_{\ell,k,n} \leq d_{\ell,k,n-1} + \Phi^{\max}_{\ell,k} p_{\ell,k,n-1} \tag{C16}$$
$$d_{\ell,k,n} \geq d_{\ell,k,n-1} - \Phi^{\max}_{\ell,k} p_{\ell,k,n} \tag{C17}$$
$$d_{\ell,k,n} \leq \Phi^{\max}_{\ell,k}(1 - p_{\ell,k,n} + p_{\ell,k,n-1}) \tag{C18}$$

Then we constrain the phase duration to be between $\Phi^{\min}_{\ell,k}$ and $\Phi^{\max}_{\ell,k}$,

$$d_{\ell,k,n} \leq \Phi^{\max}_{\ell,k} \tag{C19}$$
$$d_{\ell,k,n} \geq \Phi^{\min}_{\ell,k}(1 - p_{\ell,k,n}) \tag{C20}$$

Finally, we constrain the sum of all the phase durations for light ℓ to be within the cycle time limits $\Psi^{\min}_{\ell}$ and $\Psi^{\max}_{\ell}$,

$$d_{\ell,1,n-1} + \sum_{k=2}^{|\mathscr{P}_\ell|} d_{\ell,k,n} \leq \Psi^{\max}_{\ell} \tag{C21}$$
$$d_{\ell,1,n-1} + \sum_{k=2}^{|\mathscr{P}_\ell|} d_{\ell,k,n} \geq \Psi^{\min}_{\ell}(p_{k,1,n} - p_{k,1,n-1}) \tag{C22}$$

Note, that in C21 and C22 we use the duration of phase 1 from the previous interval, $n-1$, since when we arrive at the beginning of the next cycle of light ℓ and the phase sequence starts again from phase 1, $d_{\ell,1,n} = 0$ and $d_{\ell,1,n-1}$ is set to the duration of the previous activation of phase 1, and we can sum the total duration of the last cycle across all the phases. Additionally in C22 we activate the minimum cycle time constraint at exactly the beginning of the cycle with the signal $p_{k,1,n} - p_{k,1,n-1}$. This is illustrated in figure 1(d).

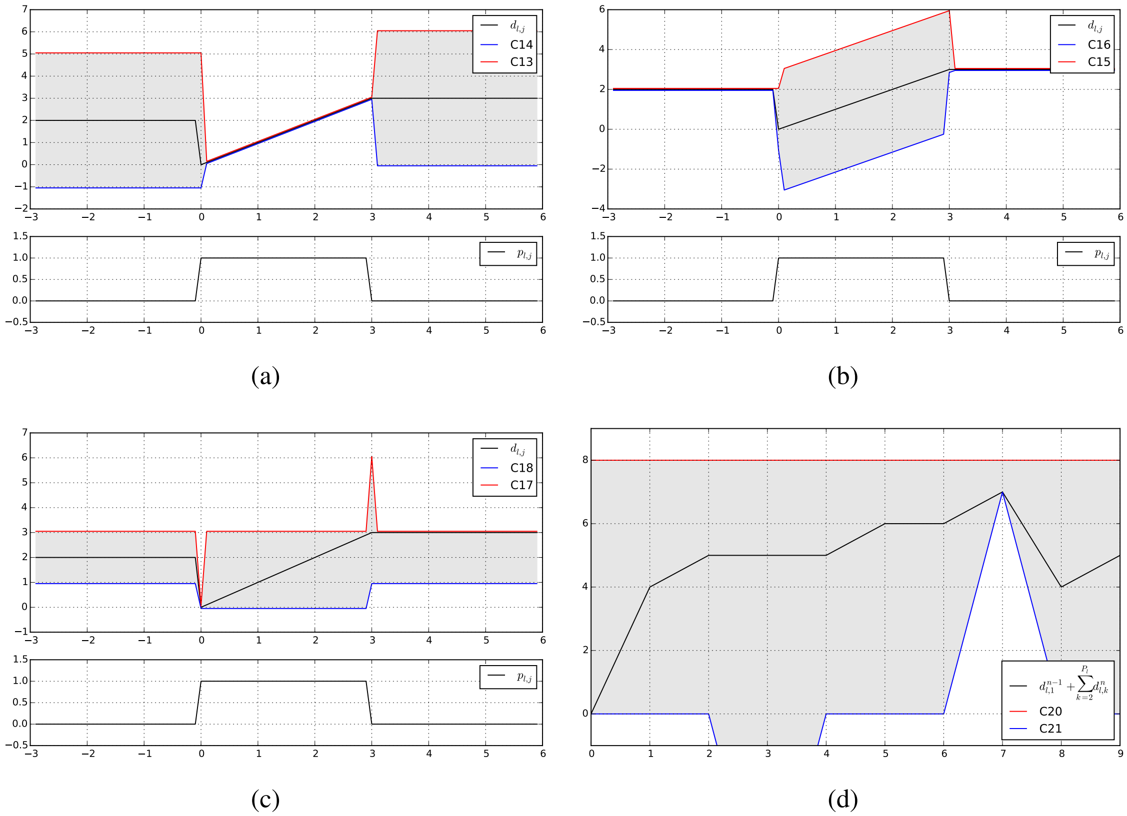


FIGURE 1 An example showing the phase and cycle time constraint envelopes. In (a), (b) and (c), $\Phi_{\ell,k}^{\min} = 1$ and $\Phi_{\ell,k}^{\max} = 3$, the duration of the previous activation was 2 and the duration of the current activation is 3. In (d), the total cycle time is 7 with $\Psi_{\ell}^{\min} = 7$, $\Psi_{\ell}^{\max} = 8$

Objective Function

Lin and Wang (1) derive an objective function for the minimisation of total delay based on the difference between the cumulative departure and arrival curves at the origin and destination. However, such an approach requires the network to be cleared at the end of the optimisation period. We derive an objective function for the maximisation of flow in the network, and apply it to every queue in the network at $q_{i,n}^{\text{out}}$,

$$maximise\left(\sum_{n=1}^{\mathrm{N}}\sum_{i=1}^{|\mathscr{Q}|}(\mathrm{T}-t_n+1)q_{i,n}^{\text{out}}+\sum_{n=1}^{\mathrm{N}}\sum_{i=1}^{|\mathscr{Q}|}(\mathrm{T}-t_n+1)in_{i,n}\right) \tag{O1}$$

And with the addition of the second $in_{i,}$ term, O1 also ensures that C1, C2, and C4 are also at their maximum upper bound.

ANALYSIS

Networks

Three networks of increasing complexity were defined for performance analysis and comparison. The first consists of a two-way avenue with three traffic light controlled intersections with 3 one-way side streets, as shown in figure 3. The second is an extension of the first to include a second parallel two-way avenue and an additional 3 traffic lights to control the side streets, as shown in figure 4. And the third is a grid of three EW two-way avenues, with two NS two-way avenues and one NS one way street and a NE to SW diagonal one-way street, as shown in figure 5. The traversal time of each queue in all three networks is set at 9 seconds (a distance of about 100m with a free flow speed of 50km/h). The maximum capacity of each queue is set at 60 cars. Flows are defined only straight ahead from the head of each queue into the tail of the next - there is no turning traffic, and in all cases the maximum flow rate is set at 5 cars per second. Each traffic light has two phases - NS and EW. The minimum phase time is 1 second and the maximum phase time is 3 seconds. The minimum cycle time of both the phases is 2 seconds and the maximum is 6. For each network a background level of flow is first established and then later increased as a wave of higher volume traffic is injected into the network at some of the streets. Then all the traffic is allowed to clear the network before ending the simulation to support an analysis of the total travel time in the network. The details of the flow levels is given in tables 3, 4 and 5.

Experiments

With a CTM based model the Δt must remain fixed through out the optimisation period, while the QTM supports non-homogeneous time steps mixed with homogeneous time steps. To exploit this we define a mutli-step solver that allows the time step of the plan to increase overtime. Such an increase will reduce the total number of variables for the same optimisation period compared to a homogeneous plan, with the trade off that the plan will be have less resolution over time. But this seems acceptable as the accuracy of the plan will also decrease overtime the further from the initial conditions. We start by generating a plan for a longer horizon using an increasing time step so that the solver has greater visibility of the impact of an earlier planning decision across a larger part of the network for the same number of variables as a fixed time step. We then keep the first part of this plan where the accuracy is highest and discard the rest, where the time steps where larger. Once the retained section of plan has be carried out, we generate another long horizon plan and repeat (see figure 2. We call the long term plan a major frame and the shorter section of a major frame that we retain, a minor frame. We use a minor frame of 10 seconds and increasing major frame sizes from 20 upwards and generate such mutlistep plans with both a homogeneous Δt of 0.25 seconds and a non-homogenous Δt ranging from fixed 0.25 second increments during the minor frame and the increasing linearly until 1 second at the horizon. We generated plans for all three networks using both the homogeneous Δt major frames and the non-homogeneous Δt major frames. For reference we also performed a full optimal solve using a fixed Δt of 0.25 seconds. Once we have generated a set of minor frames, we combined them into a large fixed plan and simulate the flow of in the network with a fixed Δt of 0.25, to support a fair comparison.

RESULTS

We compared the performance of non-homogeneous and homogeneous Δt in two ways: comparing the decrease in total travel time with increasing major frame size. And analysing the distribution of delay in each queue of the network. Figure 6 (a), (c) and (e) show a comparison between the number of time samples used in the major frame vs the % improvement in total travel time. It can be seen that using a non homogenous Δt converges towards the optimum more quickly than the homogeneous Δt for the same number of time samples. Figure 6 (b), (d) and (f) show a comparison of distribution of delay across the network. This gives us an indication of the quality of the solution in terms of the number of vehicles that experience significant delay and if the plan

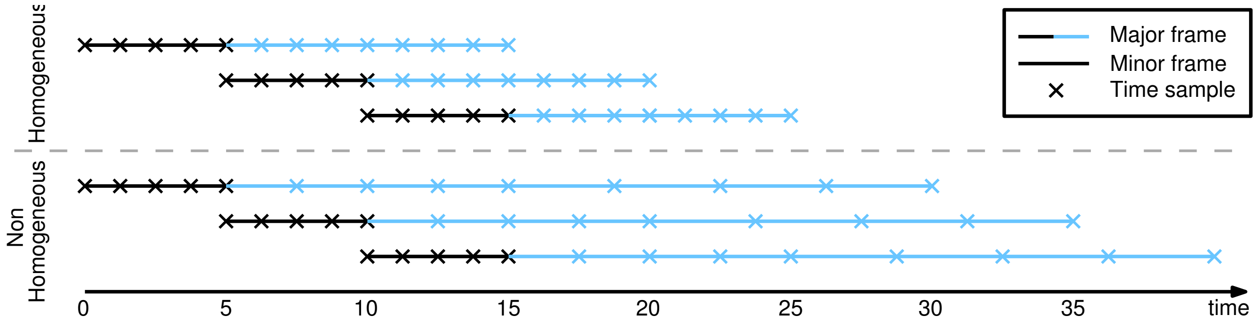


FIGURE 2 Multi-step planning

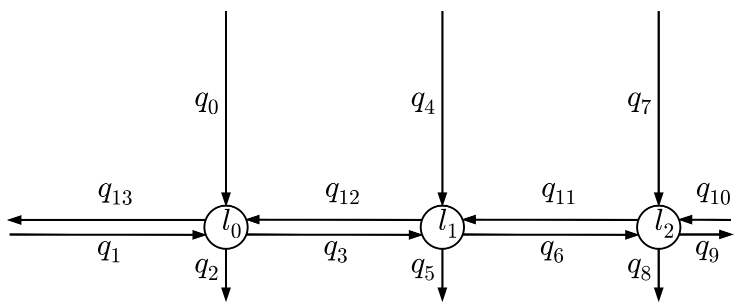


FIGURE 3 Network 1

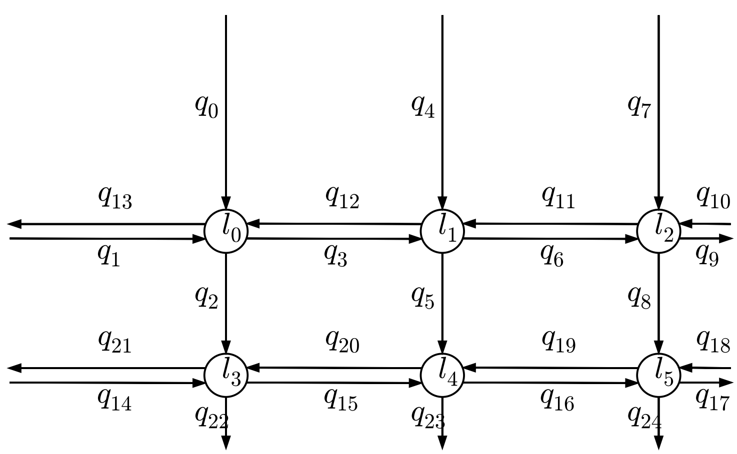


FIGURE 4 Network 2

may be starving some parts of the network. The plots show three comparisons: at the point where the non-homogeneous Δt first converges on the optimum solution, where the homogeneous Δt first converges on the optimum solution, and the optimum solution. With all three networks the quality of the solutions improves or stays the same using an non-homogeneous Δt compared to a homogeneous Δt. Finally, figure 7 shows the how cumulative arrival and departure curves and the

Queue	Background	End	Wave	Start	End
q_0	1	85	1	55	70
q_1	2	85	4	55	70
q_4	4	85	4	55	70
q_7	4	85	4	55	70
q_{10}	2	85	4	55	70

TABLE 3 Network 1 traffic parameters

Queue	Background	End	Wave	Start	End
q_0	1	85	1	55	70
q_1	2	85	4	55	70
q_4	4	85	4	55	70
q_7	4	85	4	55	70
q_{10}	2	85	4	55	70
q_{14}	4	85	4	55	70
q_{18}	2	85	4	55	70

TABLE 4 Network 2 traffic parameters

delay develop over time for q_1 of network 2. Figure 7 (a) shows the comparison at the point where the non-homogeneous Δt first converges and shows that with the longer major frame of the non-homogeneous Δt, it is able to adopt a better signal plan early on to anticipate the wave of traffic that arrives at about the 55 second point, while the homogeneous Δt with its shorter major frame initially prioritises the side street (q_0, q_2, q_{22}) over q_1 resulting in significant delay once the wave of traffic arrives. Once homogeneous Δt has converged in Figure 7 (b), both plans are close to the optimum shown in Figure 7 (c).

CONCLUSION

We have demonstrated that by exploiting the non-homogeneous time steps supported by the QTM, we are able to scale the model up to larger networks and using the same number of binary variables as a homogeneous time step, and with the same quality of a homogeneous solution using more binary variables.

REFERENCES

[1] W. Lin and C. Wang, *An Enhanced 0-1 Mixed-Integer LP Formulation for Traffic Signal Control*, IEEE Transactions on Intelligent Transport Systems, Vol. 5, No. 4, pp. 238–245, December 2004.

[2] C. F. Daganzo, *The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory*, Transport. Res. B., vol. 28, no. 4, pp. 269Ð287, 1994.

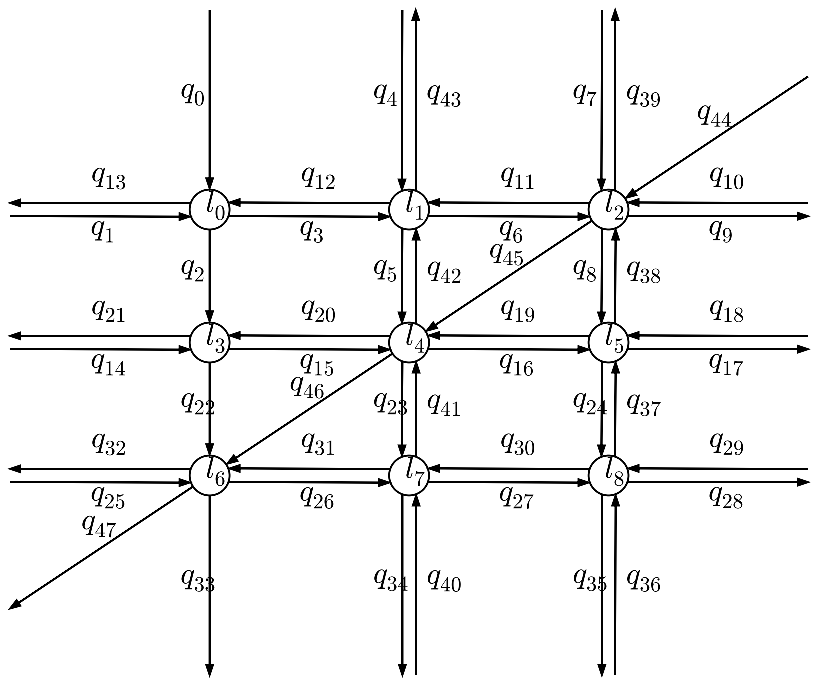


FIGURE 5 Network 3

Queue	Background	End	Wave	Start	End
q_0	1	85	1	55	70
q_1	2	85	4	55	70
q_4	4	85	4	55	70
q_7	4	85	4	55	70
q_{10}	2	85	4	55	70
q_{14}	4	85	4	55	70
q_{18}	2	85	4	55	70
q_{25}	4	85	4	55	70
q_{29}	2	85	4	55	70
q_{36}	2	85	4	55	70
q_{40}	2	85	4	55	70
q_{44}	2	85	4	55	70

TABLE 5 Network 3 traffic parameters

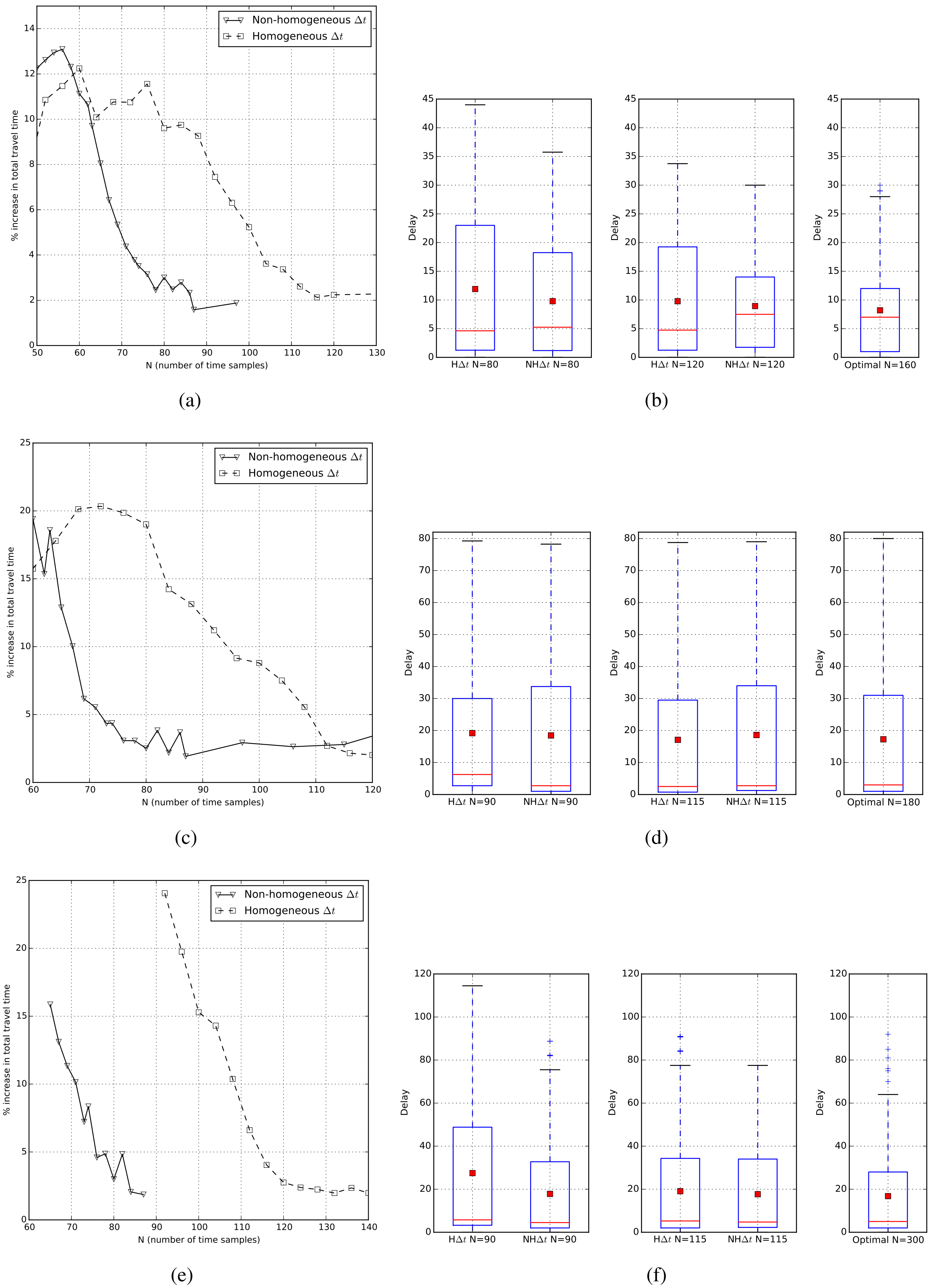
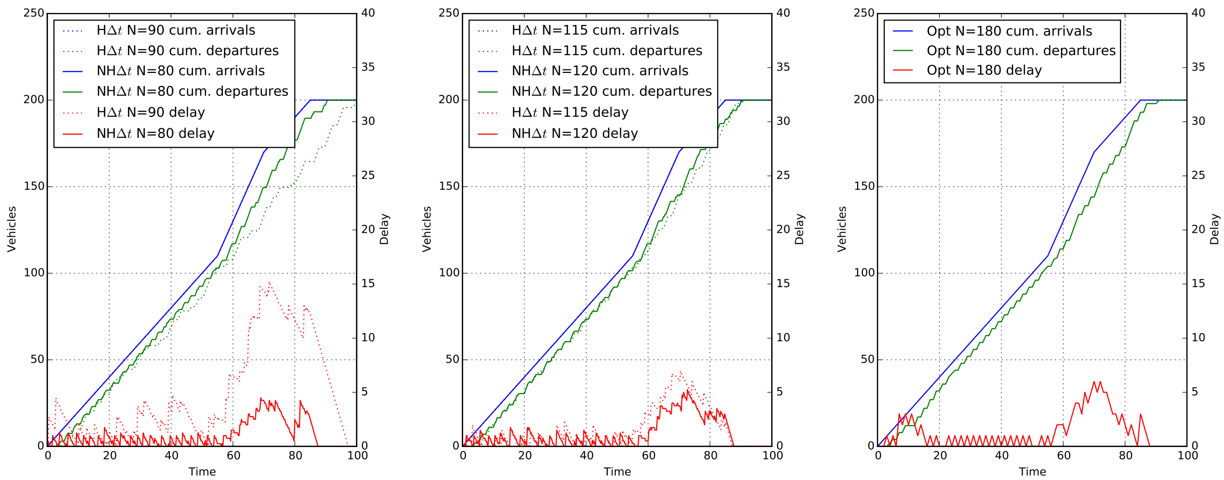


FIGURE 6 Results for the three networks showing the comparitive % improvement in total travel time for the network between using a homogeneous Δt and a non-homogeneous Δt, and the distribution of delay time at the convergence point of non-homogeneous Δt, the convergence point of homogeneous Δt and for the fully solved optimal solution. (a) and (b) 3 light avenue, (c) and (d) 6 light grid, and (e) and (f) 9 light grid,



(a) (b) (c)

FIGURE 7 Cumulative arrival and departure curves and delay for queue 1 in the 6 light grid. (a) at the convergence point of the non-homogeneous Δt it is near to the optimum solution while homogeneous Δt lags behind (b) at the convergence point of homogeneous Δt both are near optimum, and (c) the fully solved optimal solution