NISS

Direct Signal Timing Optimization: Strategy Development and Results

Nagui M. Rouphail, Byungkyu "Brian" Park, and Jerome Sacks

Technical Report Number 109
November, 2000

National Institute of Statistical Sciences
19 T. W. Alexander Drive
PO Box 14006
Research Triangle Park, NC 27709-4006
www.niss.org

DIRECT SIGNAL TIMING OPTIMIZATION:
STRATEGY DEVELOPMENT AND RESULTS

Nagui M. Rouphail, Professor, Department of Civil Engineering
North Carolina State University, Raleigh, NC, USA
rouphail@eos.ncsu.edu
Byungkyu "Brian" Park, Post-Doctoral Fellow, National Institute of Statistical Sciences
Research Triangle Park, NC, USA
park@niss.org
Jerome Sacks, Senior Fellow, National Institute of Statistical Sciences
Research Triangle Park, NC, USA
sacks@niss.org

ABSTRACT

Direct optimization pertains to the use of a single traffic model both for signal timing generation and for plan evaluation. Criteria for model selection include realistic traffic representation, adequate breadth to incorporate most urban traffic management features, and the ability to represent system variability. In the U.S., the CORSIM model is the closest model meeting these requirements.

A small urban traffic network, with nine signalized intersections in the city of Chicago, U.S.A was tested. Extensive field studies gathered all the inputs for the U.S. version of TRANSYT-7F (T7F) and CORSIM simulation model in the AM and PM peak hours. Field measurements under this base case show that the queue lengths produced by CORSIM are similar to the field values.

Traditional signal optimization was carried out using the T7F package. Twelve different signal strategies were tested in T7F and the one giving the best overall performance in CORSIM was selected. Each evaluation involved 100 confirmation runs in CORSIM. Two measures of effectiveness (MOE's) were used: link delay and total network queue time. The mean, median and standard deviation for each MOE were produced for each set of the 100 confirmation runs.

Direct signal optimization was performed using a genetic algorithm (GA). GA is a guided random search that uses the concepts of natural section and evolution to evaluate and propose improved solutions by optimizing a given objective. Given adequate computing resources, GA converges to an optimal (not necessarily global) solution. A difficulty in CORSIM application is the inherent variability in system output, which slows down convergence.

The results indicate that overall network performance improves dramatically under the direct-optimization, GA-determined settings. Both mean and median MOE values were substantially lower than T7F and the base case, as was the variability in system performance.

KEYWORDS
Signal, Optimization, Genetic Algorithm

1. INTRODUCTION

The development of efficient signal timing plans for urban traffic networks has always been a challenging task for the traffic analyst. These networks can be quite complex in nature, serving a variety of vehicular and non-motorized users, and private as well as public transportation modes. Further, the performance of signal control strategies on such networks is quite difficult to predict due to the stochastic nature of traffic flows, as evident by day-to-day variations in traffic demand, vehicle composition and service times. By extension, the production of signal control strategies that can effectively respond to such variations is also quite difficult to achieve. It is no coincidence, therefore, that signal timing methods are developed almost exclusively in macroscopic, deterministic traffic environments. For example, all the traditional optimization models for isolated, signalized intersections cited in Click and Rouphail (1999) fall in the category of macroscopic deterministic approaches.

Direct optimization in the context of this document refers to the use of a single, high-fidelity traffic model both for signal timing generation and for plan evaluation. Direct optimization provides a highly flexible environment for solving the signal timing optimization problem. Any measure (or combination of measures) of effectiveness produced by the model can be used. Link-based or network wide constraints can be incorporated, so can advanced signal control logic such as the designation of subnetworks and double cycling. Finally, time-dependent signal settings can be derived, as long as time-dependent demands can be accommodated in the model.

The criteria for model selection include an ability to produce a realistic representation of the traffic environment, adequate model breadth to incorporate most urban traffic management features (e.g., parking, STOP control, bus stops and routes) and an ability to represent system variability both in time and space. In the U.S.A., the microscopic, stochastic CORSIM (1997) model is the closest one to meeting all these requirements. Further, CORSIM has a long history of acceptance by traffic professionals and support from the U.S.A Federal Highway Administration and State Departments of Transportation. Variations of this model have been used in the U.S. for over thirty years. While recognized as an excellent traffic simulator, CORSIM has no optimization capabilities. Therefore, an optimization interface with CORSIM is required in order to enable direct optimization.

1.1 Organization

The paper is organized as follows. Section 1 concludes with a listing of the study objectives and a review of the technical literature pertaining to traditional and direct optimization methods. Section 2 discusses the study design including network description, data collection and model verification. Section 3 describes how the signal-timing problem is formulated, using both traditional and direct optimization methods. Section 4 provides a summary of the findings, and compares the performance of various methods. Conclusions and recommendations for further research are given in Section 5.

1.2 Study Objectives

The principal objectives of this study were twofold:

- to test the performance of traditional network signal optimization strategies in microscopic, stochastic traffic environments.
- to develop and test search methods that can be used for direct traffic signal optimization.

1.3 Literature Review

The state of practice in traffic signal timing in the U.S. has not appreciably changed over the past twenty years, although research and development efforts in the area of adaptive signal control (Gartner and Stamatiadis, 1998) have greatly accelerated. For a review and field assessment of signal optimization software at isolated intersections, the reader is referred to Click and Rouphail (1999). At the arterial or network level models such as PASSER-II (Messer et. al 1990), MAXBAND (Little et. al 1980), TRANSYT-7F—or T7F (Wallace et. al 1998) and SYNCHRO (Trafficware 1999) have been used. It is emphasized that "T7F" refers to the U.S. Federal Highway version of the TRANSYT model that has been developed and updated in the U.S. since the mid 1970's, and is <u>not</u> related to any recent TRANSYT modeling activities in the U.K.

T7F uses cyclic flow profiles (CFP) to project and disperse traffic on the links, while SYNCHRO uses average flow rates to predict cycle-average traffic performance. Both models can generate optimal plans for a network, including the specification of a system-wide cycle length, movement green times, and intersection offsets. SYNCHRO uses an exhaustive search to determine optimal fixed-time signal plans, while T7F (because of its more complex traffic model) applies a hill-climbing heuristic approach to accomplish the same objectives. The latest release of T7F (known as Release 8) is designed to simulate and produce good signal settings under congested traffic flow conditions including spillback.

A recent study conducted by Park et al. (2000) found that the reliability of T7F (based on CORSIM evaluations) was not very satisfactory for the test network used. They reported that the selected system performance measures in T7F did not match well to those resulting from CORSIM runs. They concluded that this is due to the use of a low fidelity traffic model during T7F optimization, and recommended the use of a more realistic simulation program that can account for the stochastic nature of traffic demand and driver's responses. Park et al. (1999) also found limitations to T7F for over-saturated conditions. They pointed out that T7F tended to produce higher cycle lengths to reduce random-plus-oversaturation delay by increasing the apparent phase capacity. This eventually resulted in more chances of spillback and blockage especially for closely spaced intersections. Interestingly, similar concerns regarding T7F limitations were voiced when it was used as the "base case" optimization to be compared with a proposed adaptive signal control strategy (Tarnoff and Gartner, 1993).

Direct optimization in the context of the CORSIM model environment can be best described as a stochastic optimization (SO) problem. SO algorithms have been applied in the areas of plant locations and shipment (LeBlanc, 1977; Franca et. al, 1982), and for transportation network assignment (Sheffi, 1985). However, no significant studies applying SO principles to the area of traffic signal control were found. The class of methods known as genetic algorithm (GA) offers one route to address this problem.

GA is a guided random search technique widely used in optimization (Goldberg, 1989; Rudolph, 1994; Cerf, 1995). Since GA makes no assumptions about the traffic environment, but can interact with CORSIM to produce optimal plans based on CORSIM's predicted system performance, it was deemed to be the method of choice. Implementing GA in connection with CORSIM required an interface between two the programs to enable the full functionality of the direct optimization feature. These features are described in Section 3 below.

2. FIELD STUDY DESIGN

This section describes the test site, the data that were collected at the site, and the initial CORSIM model verification.

2.1 Site and Data Collection Description

The test network used in this study is located in the City of Chicago in the State of Illinois, USA. A schematic of the site is depicted in Figure 1. The network consists of 9 signalized intersections, 59 one-way links and 31 internal nodes. Most intersections operate under two-phase signal control with permitted left turns. The data collection effort was designed to provide both CORSIM input parameters and output measures of performance. On the input side, traffic counts for vehicles as well as pedestrians were collected. Vehicle arrival rates at the boundary nodes were measured for one hour, and turning volumes were collected at all intersections. For key intersections only, all vehicle flows were measured for a full hour. Bus data were also collected, as well as all link geometry features (free flow speed; number of lanes; grades, etc.). Finally, all signal timing parameters including cycle length, phase times and offsets were measured. These data constituted the "base case" settings against which various optimization schemes were contrasted. On the output side, the maximum queue length (MQL) at key intersections was observed during the one-hour observation period and recorded by manual observers. Data were collected during weekday AM and PM peak hours.

2.2 CORSIM Model Coding and Verification

The test network was coded in a CORSIM input data file. CORSIM's default values were used for the distributions of driver types, spillback probabilities, queue discharge headways, and many other inputs. The measured inputs were: external input demand, turning percentages, traffic mix, bus routes, bus headways, bus dwell times, basic geometry, posted speed limit, signal timing plan. No changes in the phase patterns or left turn treatments were considered (neither in the base case nor in the optimization phase). CORSIM's performance was tested by comparing the field MQL to the distribution of MQLs generated by repeated CORSIM runs using different random number seeds. One hundred CORSIM runs were executed in order to account for the stochastic variability in traffic demand and driver behavior. As shown in Figure 2, the field MQLs are consistent with the MQL distributions for the Northbound and Southbound through movements at the intersection of Western and Lawrence (Figure 1). Further details on the field data collection and network evaluation can be found in Park et al. (2000).

3. FORMULATION OF OPTIMAL SIGNAL TIMING PLANS

3.1 Traditional Method

Traditional signal optimization was carried out on the study network using a calibrated T7F model. Calibrated model parameters included link saturation flow rate, left turn "sneakers" and link free flow speed. Of course, all input parameters were entered directly based on field observations. Twelve separate signal strategies were tested in T7F. Among these were minimum delay, disutility index or DI, fuel consumption, and progression opportunities or PROS. Each strategy was evaluated using 100 CORSIM runs, resulting in a total of 1,200 simulations. Network performance was gauged by link delay and total network queue time. The mean, median and standard deviation for each MOE were calculated for each set of 100 confirmation runs. The strategy yielding the best overall network performance when implemented in CORSIM (version 4.2) was subsequently selected as the T7F strategy. For the test network, the best T7F strategy was the maximization of $(PROS/DI)^{100}$. For additional details on the T7F optimization, the reader is referred to Hochanadel (1999).

3.2 Direct Optimization

The direct optimization process is carried out in the following manner. First, a Rexx code (QUECUS 1996) was developed to act as the interface between CORSIM and the GA optimizer. It interacts with the CORSIM input file (.TRF) by inserting experimental values of a network signal plan (cycle, splits and offsets), as well as a variable random number seed in each run. After a CORSIM run is completed, the code interacts with the output file (.OUT) by extracting the relevant set of link and system performance measures from it and routing these to the optimizer. The GA optimizer uses these values, and updates the experimental input signal plan values for the next series of experiments. The process continues until a predetermined maximum number of generations are reached. For the network depicted in Figure 1, there were 22 decision variables (one system cycle length, seven offsets and fourteen independent green splits). The GA optimizer parameters are: population size = 25; maximum number of generations = 25; uniform cross over probability = 0.40 and mutation probability = 0.03. These parameters were found to be adequate after examining a few smaller problems with fixed cycle lengths and splits. The objective function in the direct optimization scheme is the modified network queue time (*MNQT)* defined by:

$$MNQT = \sum_{i=1}^{L} \left[QT(i) \times \left\{ \max\left(1, \frac{MQL(i)}{SC(i)} \right) \right\} \right] \ldots\ldots\ldots\ldots\ldots (1)$$

where: *QT(i)* = queue time on link i, i= 1,2,…., L (=59)
MQL(i) = maximum queue length on link i observed during the simulation
SC(i) = through signal capacity on link i, calculated as the product of the link discharge rate and the green to cycle ratio for the link.

The objective function in equation (1) inhibits the formation of very long queues on links, even though adequate queuing space (reflected in high values of MQL) may be available. This is accomplished through the introduction of the penalty term in curly brackets. The solution constraints include the specification of minimum green times to accommodate pedestrian crossings, and offsets that can vary from zero to (cycle length-1) seconds. Of course, all change intervals were kept fixed in all simulations.

One rationale for choosing network queue time instead of delay as the objective function is the limitation of CORSIM's reporting of delay. Link delay is reported by CORSIM only for those vehicles that have crossed the link during the simulation period. On the other hand, queue time represents a cumulative measure of all queuing that occurs on the network, and includes the delay experienced by vehicles that remain in the network, or those that have yet to cross a link at the end of the simulation period.

4. RESULTS

In this section, we present the results obtained from direct optimization and compare the results with those from some traditional signal control strategies.

4.1 GA Performance

In an initial set of GA runs, a random starting point was used and green splits were only bounded by the minimum green times. Convergence of the GA algorithm was inadequate for the GA parameters tested. We reduced the search space by limiting the green splits to be within ± 0.15 of the "equal v/c ratio" solution as recommended by Webster (1958). The result was a vast improvement in the convergence, and a reduction in average delay from 32 to 15 seconds/vehicle. A graphical representation of the GA convergence properties for the average and minimum delay in each generation is depicted in Figure 3. In this graph, it is clear that convergence occurs after about twenty to twenty-five generations. On the other hand, the minimum delay solution fluctuates between 13s and 18s.

The optimization of the signal parameters using 25 generations and population size of 25, required 625 CORSIM runs and consumed about 7.8 hours (or 45 seconds per run) on a single Pentium II (400 MHz) machine. In addition, one hundred CORSIM confirmation runs were used to evaluate the distribution of traffic performance. These confirmation runs consumed an additional hour (or about 30 seconds per run).

4.2 Traffic Performance

Comparisons of traffic performance using traditional and direct signal optimization methods were carried out for each of two peak hours where data were available. Both network delay (the aggregation of link delays) and network queue time (see Section 3.2) were evaluated. Since repeated observations were made using one hundred confirmation runs, the distributions of the performance measures could be compared. Table 1 is based on results obtained using the latest CORSIM version (4.32), and summarizes the study findings. For each time period, and for each of three signal strategies (Base, T7F and GA), the table gives the best cycle length and the mean, standard deviation and median for vehicle delay and network queue time based on 100 runs.

It is evident from these results that the GA settings outperform T7F as well as the base settings, in both AM and PM peak hours. Compared to the base setting, the GA settings reduced the simulated delays from 30-44% and the simulated queue times from 44-61%. The GA results were even more impressive when compared to the best T7F settings. Note that the GA produced slightly higher cycle lengths than T7F, which may in part explain the delay differences. Further investigation at the link level indicates that T7F appears to have optimized the performance of North-South traffic on Western Avenue at the expense of cross street traffic. GA, on the other hand, prevented the occurrence of long queues anywhere on the network as a consequence of the large penalties associated with such queues (see Section 3.2). Other differences between the strategies might be attributed to the different ways in which T7F and CORSIM represent traffic, and therefore their own view of "optimality". But, when the T7F settings are selected on the basis of their performance in CORSIM, they are still inferior to the GA settings.

The results also indicate that good signal settings decrease variability as well as improve average system performance. This is particularly evident in Figure 4, where distributions are compared. Both T7F and base settings result in a much wider range of queue time than GA. The same observations pertain to the results obtained for the PM peak. Of course, the ultimate test of effectiveness must be through a field evaluation of the proposed plans.

5. CONCLUSIONS AND RECOMMENDATIONS

This paper presented the application of genetic algorithms to the development and evaluation of optimal signal timing settings for a real-world urban traffic network. The development phase encompassed the preparation of computer codes to interface with CORSIM input and output files, and the testing of a variety of model parameters and solution spaces. That phase of the work produced a direct optimization method that is applicable to small size networks, similar to the one depicted in Figure 1. The evaluation phase included collecting empirical data for model input and evaluation, and a comparison of the signal setting derived from direct optimization with those obtained from traditional methods (e.g. TRANSYT-7F). Comparisons were made on the basis of repeated observations of the CORSIM traffic model.

We found that the GA settings consistently outperform TRANSYT-7F strategies, even though the “best” T7F strategy was selected on the basis of how it performed in CORSIM, not in TRANSYT-7F. The results are consistent for the peak periods and performance measures studied. The best strategy not only improves the mean value of the performance measure, but also reduces its variance. These findings have implications on the way benefits of advanced signal control strategies are estimated, since (for the most part) these are based on the use of T7F as the optimum “before” case.

Although successful, our experience also points to the need for further work in the application of GA optimization methods to signal control. In particular, scalability (to larger or more congested) networks needs to be addressed in a systematic way. Methods to reduce or control the variance in the simulated experiments in order to expedite convergence should also be explored. Last but not least, field verification of the GA settings is essential in order to provide credibility to the simulation results. Plans are underway to carry out a field evaluation on a similar network in Chicago.

Acknowledgements

This research was sponsored by NSF grants DMS-931303 and DMS-9208758. The authors thank Dr. Vonu Thakuriah for advice and comments on this research and for directing the collection of data by the Urban Transportation Center at the University of Illinois at Chicago used in this research.

References

Cerf, R. (1995) An Asymptotic Theory for Genetic Algorithms, Artificial Evolution, Lecture Notes in Computer Science 1063, Springer, 37-52.

Click, S. and Rouphail, N. (1999) Lane Group Level Field Evaluation of Computer-based Signal Timing Models, Paper presented at the 78th Annual Meeting, TRB.

CORSIM User's Manual. (1997), **FHWA, U.S. Department of Transportation**, Office of Safety and Traffic Operation R&D, Intelligent Systems and Technology Division, McLean, VA.

Franca, P. M. and Luna, H. P. L. (1982) Solving Stochastic Transportation-Location Problems by Generalized Benders Decomposition, Transportation Science 16, 113-126.

Gartner, N. H., Stamatiadis, C., and Tarnoff, P. J. (1995) Development of Advanced Traffic Signal Control Strategies for Intelligent Transportation Systems: Multilevel Design, Transportation Research Record 1494, TRB.

Goldberg, D. E. (1989). Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley, Reading.

Hochanadel, J. (1999) Statistical Evaluation of Alternate Signal Timing Plans, Master's Thesis, North Carolina State University, Raleigh, NC, USA.

Le Blanc, L. J., et al. (1977) Heuristic Approach for Large Scale Discrete Stochastic Transportation-Location Problems, Comp. and Math. with Applications 3, 87-94.

Little, J. D. C., and Kelson, M. D. (1980) Optimal Signal Timing for Arterial Signal Systems, MAXBAND. Operations Research Center, MIT, Cambridge, MA, USA.

Messer, C. J., Haenel H. E. and Koppe E. (1974) A Report on the User's Manual for Progression Analysis and Signal System Evaluation Routine – PASSER II. TTI Research Report 165-14, College Station, Texas.

Park, B., Messer, C. J., and Urbanik II, T. (1999) Traffic Signal Optimization Program for Oversaturated Conditions: Genetic Algorithm Approach, Transportation Research Record 1683, 133-142.

Park, B., Rouphail, N. M., Hochanadel J., and Sacks, J. (2000) Evaluating the Reliability of TRANSYT-7F Optimization Schemes, Paper presented at the 79th Annual Meeting, TRB.

QUECUS Systems (1996) Personal REXX User's Guide.

Rudolph, G. (1994) Convergence Analysis of Canonical Genetic Algorithms, IEEE Transactions on Neural Networks 5, 96-101.

Sheffi, Y. (1985) Urban Transportation Networks, Prentice-Hall Inc., New Jersey.

Tarnoff, P. J. and Gartner, N. (1993) Real-Time Traffic Adaptive Signal Control, Proceedings of the Advanced Traffic Management Conference: Large Urban Systems, 157-168.

Trafficware (1999) Synchro 4.0 User's Guide, Albany, CA, USA.

Wallace, C. E., Courage, K. G., Hadi, M. A., and Gan, A. C. (1998) TRANSYT-7F User's Guide. Transportation Research Center, University of Florida, Gainesville, FL, USA.

Webster, F. V. (1958) Traffic Signal Settings. Road Research Technical Paper, No. 39, Her Majesty's Stationary Office, London, U.K.

Table 1. Comparative performance of signal strategies in CORSIM

Time Period	Signal Strategy	Cycle Length	Delay (sec/vehicle)			Queue Time (veh-hrs)		
			Mean	S.dev.	Median	Mean	S.dev.	Median
AM Peak	**Base**	65,85	27.8	3.8	26.3	232	38	218
	Best-T7F	65	33.1	6.3	33.2	304	56	310
	GA	75	19.5	0.30	19.5	130	13	128
PM Peak	**Base**	65,85	29.8	6.1	29.3	299	94	284
	Best-T7F	65	26.5	5.2	25.1	378	124	371
	GA	70	16.4	1.4	16.2	116	14	113

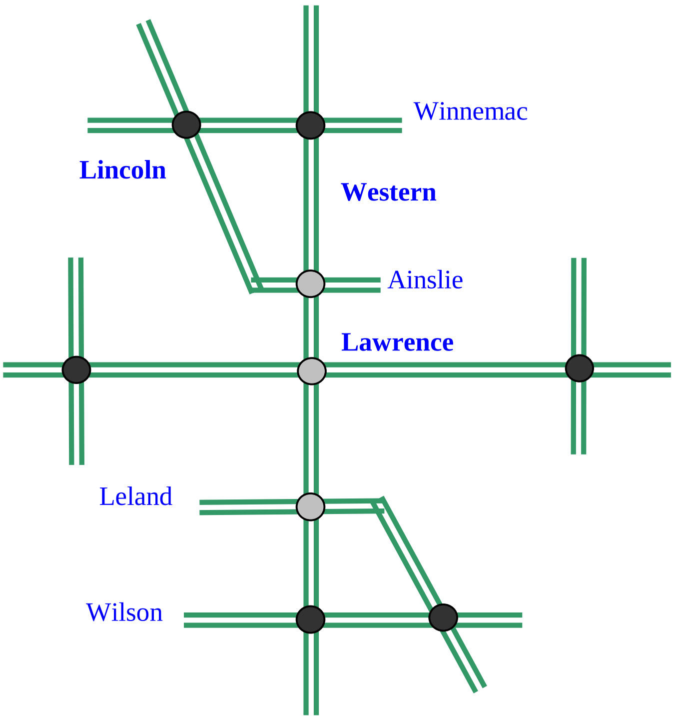
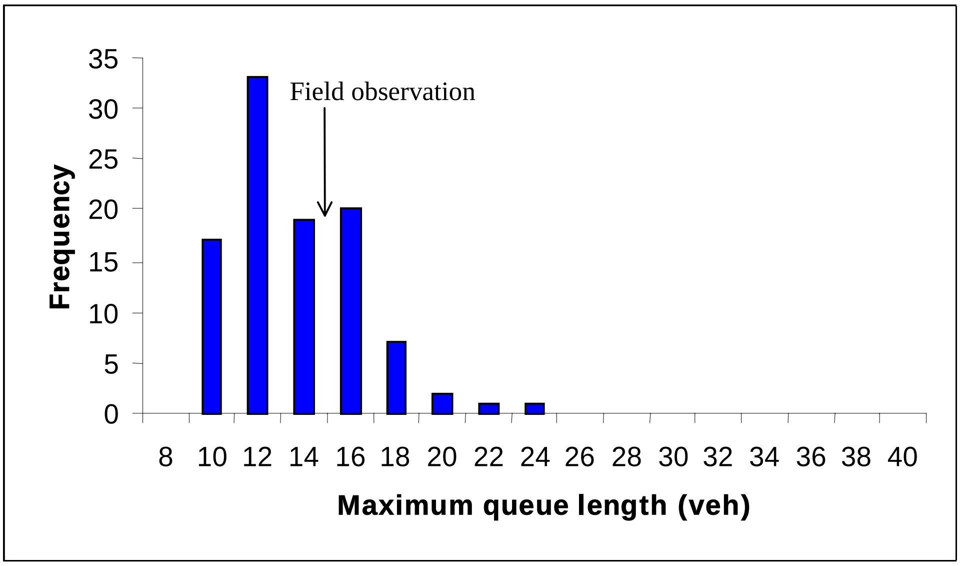
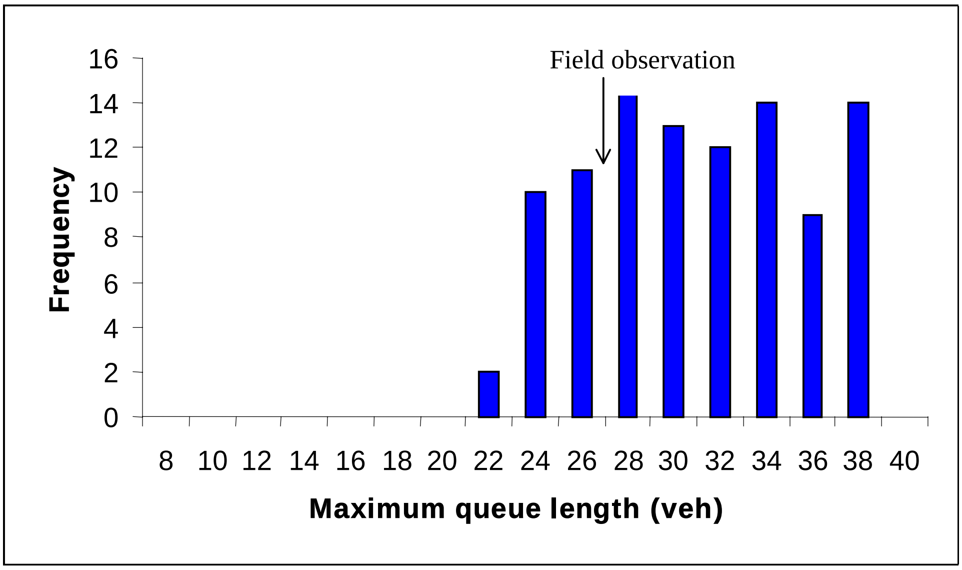


Figure 1. Schematic of the test network



(a) Southbound (100 CORSIM simulations)



(b) Northbound (100 CORSIM simulations)

Figure 2. MQL distribution from CORSIM and field at intersection of Western and Lawrence with base signal timing plan

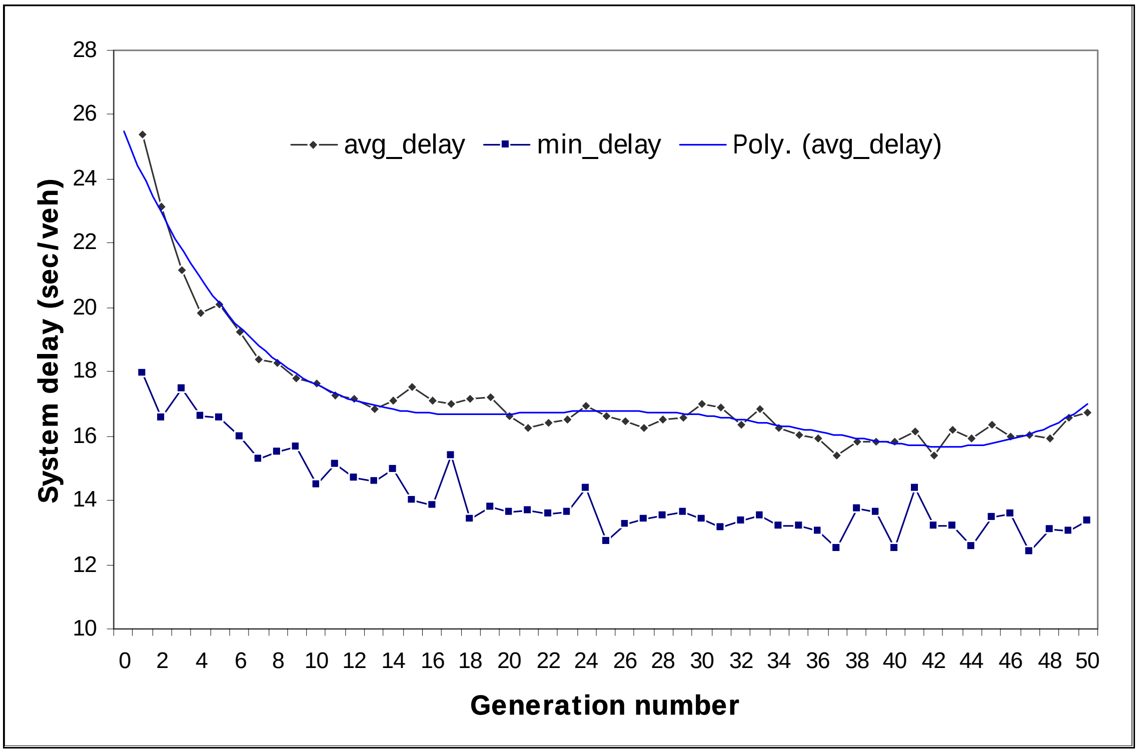


Figure 3. Convergence of genetic algorithm

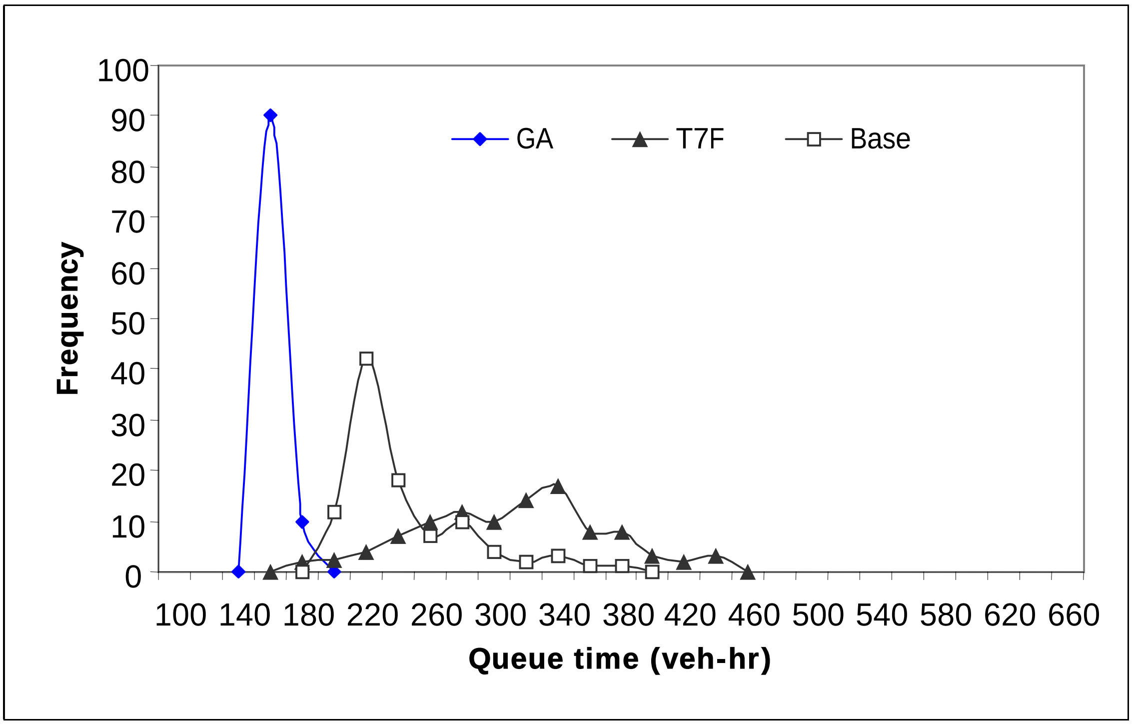


Figure 4. Comparative distribution of queue time for the AM peak period