Locating changeable message signs for advanced traffic information and management systems

Liping Fu, Jeffrey Henderson, and Shuo Li

Abstract: This paper presents an optimization model for locating changeable message signs (CMSs) on an integrated freeway-arterial network. Compared with existing models, the proposed model represents a well-balanced compromise between computational efficiency required to solve problems of realistic size, and model realism to ensure the quality of solutions. The model has three unique features: (1) it recognizes that locating CMSs is a planning problem that must take into account both current and future needs and benefits, (2) it evaluates benefits of CMSs over multiple time periods with different traffic distributions, and (3) it explicitly considers inherent variations in incident characteristics across links and over time. A sensitivity analysis is performed to examine the potential impacts on optimal CMSs locations resulting from uncertainties in various input parameters, such as traffic demand, incident attributes, and driver behaviour. Lastly, the proposed model is applied to the Highway 401 express-collector freeway system in Toronto for relocating the existing CMSs.

Key words: changeable message signs (CMSs), location optimization, traffic assignment, queuing theory.

Résumé : Cet article présente un modèle d'optimisation de l'emplacement des panneaux à messages variables (« CMSs ») sur un réseau général autoroutier. Par rapports aux modèles existants, le modèle proposé représente un compromis équilibré entre l'efficacité computationnelle requise pour résoudre des problèmes d'une dimension réaliste et le réalisme du modèle qui permet d'assurer la qualité des solutions. Le modèle présente trois caractéristiques uniques : (1) il reconnaît que l'emplacement des « CMSs » est une question de planification qui doit tenir compte des besoins et des avantages actuels et futurs, (2) il évalue les avantages des « CMSs » sur plusieurs périodes et pour différentes répartitions de trafic, et (3) il tient compte des variations inhérentes aux caractéristiques des incidents quant aux liaisons routières et dans le temps. Une analyse de sensibilité est effectuée pour étudier les incidences potentielles des incertitudes de divers paramètres d'entrée tels que la demande de trafic, les attributs des incidents et le comportement des conducteurs sur l'emplacement optimal des « CMSs ». Finalement, le modèle proposé est mis en œuvre sur le système d'artères collectrices de l'autoroute 401 à Toronto pour changer l'emplacement des « CMSs » existants.

Mots-clés : panneaux à messages variables (« CMSs »), optimization de l'emplacement, affectation de trafic, théorie des files d'attente.

[Traduit par la Rédaction]

Introduction

Changeable message signs (CMSs), also known as variable message signs (VMS), are becoming popular as one of the primary means for transportation agencies to disseminate travel and traffic information to motorists. Under the umbrella of intelligent transportation systems (ITS), CMSs constitute a key element in dynamic traffic management and information provision functions. The CMSs are commonly used to inform motorists of varying traffic, roadway, and environmental conditions and provide information on the location and severity of incidents and the expected delay. They can also be used to advise motorists of alternate routes in the event of an incident, construction, or a roadway closure.

The effectiveness of CMSs, however, depends on how many CMSs are installed and where the CMSs are located in the network. Theoretically, the benefits from CMSs can be maximized, if the whole network of roads is instrumented with CMSs. This is, however, practically impossible because of the high costs of installing CMSs. Furthermore, past studies have suggested that excessive use of CMSs could lead to diminishing returns in benefits, and even worse, could be counter-effective because of the behavioural response of drivers to real-time information (Wardman et al. 1997).

The planning of locations of CMSs is challenged by a variety of issues, such as how to model the response of drivers to changeable message sign (CMS) messages, how to model random incidents that vary by time and space, and how to model the impacts of incidents on traffic. The state-of-art practice has mostly relied on the experience and judgement

Received 9 December 2005. Revision accepted 3 October 2006. Published on the NRC Research Press Web site at cjce.nrc.ca on 26 June 2007.

L. Fu[1] and J. Henderson. Department of Civil Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada.
S. Li. College of Civil Engineering, Hunan University, Changsha, Hunan, People's Republic of China.

Written discussion of this article is welcomed and will be received by the Editor until 30 September 2007.

[1]Corresponding author (e-mail: lfu@uwaterloo.ca).

 doi:10.1139/L06-151

of planners, and thus is not the result of any comprehensive and systematic analysis. Abbas and McCoy[2] were the first in literature to study the problem of optimizing CMSs locations in a road network. Their location optimization objective was to maximize the potential reduction in vehicle delay because of traffic diversion to alternative routes in response to incident information provided by CMSs. A simple deterministic queuing model was used to estimate delays with and without CMSs in a linear freeway network. However, it was not clear how issues, such as, congestion on alternative routes, over-saturated conditions, incident rates on individual links, and dependency of diversion rate on potential savings were handled in their model.

Another relevant study was initiated by Chiu et al.[3], who proposed a bilevel stochastic integer programming model for the CMS-location problem. The location optimization problem was realized at the upper level to maximize the total user benefit of real-time information from CMSs. The responses (route choices) of users to incident conditions and information were represented at the lower level as a user-optimal-dynamic traffic assignment problem. The expected total user benefit, corresponding to a given location solution, was calculated based on a sample of benefits. Each benefit was estimated by generating a random incident on a network link and solving the resulting dynamic traffic assignment problem. The model also suffers from several limitations. First, as acknowledged by the authors, the whole process is extremely computationally intensive because of the need to evaluate a large number of candidate location plans, consider sufficient number of incident realizations for each location solution, and perform a simulation-based dynamic traffic assignment procedure for each incident realization. Second, their location benefit model was based on a route choice assumption that all users have perfect knowledge and real-time information on the network and incident conditions and possess the ability to anticipate choice of routes of other users and choose their optimal routes accordingly. Finally, it is unclear if it is practical or necessary to apply such a complex model, seemingly designed for operational management and control purposes, for solving the CMS location problem, which is essentially a planning problem.

In this paper, we propose a model that is aimed to strike a balance between computational efficiency and model complexity. The major contribution of our research is in extending the work of Abbas and McCoy[2] in three important aspects. First, we explicitly consider time-of-day variation in travel demand distribution by introducing a multi-period benefit estimation model. Second, we incorporate a logit route choice model in determining time-dependent division rate under incident conditions. Lastly, the proposed model explicitly takes into account inherent variations in incident characteristics across links and over time, such as, incident rate, incident duration,and capacity reduction.

The paper is organized as follows. We first describe the individual components of the proposed model including (*i*) a time-dependent queuing model for estimating user delay with and without the presence of a CMS, (*ii*) a dynamic traffic diversion model that relates the probability for a vehicle to divert from the incident link to its potential travel time savings, and (*iii*) a sequential optimization model for identifying the best locations for a given number of CMSs. A sensitivity analysis then follows to examine the impacts of inherent variations in input parameters on the optimal locations of CMSs.

Problem description and formulation

The CMS-location problem can be loosely defined as follows. Given a road network consisting of a set of road segments or links, identify a subset of links for installing a given number of CMSs so that the total benefit that could be obtained from these installed CMSs is maximized. While this definition is easy to comprehend, its formulation requires a formal specification of the objective measure or location optimization criterion — total benefit.

As discussed in Introduction, CMS is a media through which traffic and other relevant information can be timely delivered to drivers for improved safety and reduced travel time. As a result, the potential benefits that can be obtained from CMSs also depend on the type of information delivered by a CMS. In this research, we assume that CMS is mainly used for disseminating incident related information, and therefore its benefits can be measured by the expected total reduction in user delay or total travel time savings (TTS), which are due to traffic diversion from the congested incident locations to less congested alternative routes induced by incident and (or) route guidance information from the CMS. Note that this assumption is reasonable because incidents are the major source of traffic delay (Shrank and Lomax 2002). Other types of CMS benefits related to recurrent congestion, scheduled lane and (or) capacity reductions (e.g., construction), special events, and weather conditions are not considered in this research, but these can be easily incorporated into the proposed model once relevant benefit models are available.

To formulate the travel time savings, a number of issues still need to be addressed, such as, (*i*) how to consider the spatial and temporal variation of traffic (*ii*) how to model the uncertainty of incident occurrences (e.g., where, when, and how severe), and (*iii*) how to estimate who will pass the links with CMSs and the incident links and at what time, and who will divert. To address these issues, the following assumptions are introduced:

- A typical average day over the planning horizon is selected in calculating the total travel time savings. The analysis day is divided into different periods with each period having a constant traffic demand. It is further assumed that traffic demand by period over the area of interest, in the form of origin–destination (O–D) matrix, is available from some existing transportation planning models.
- A representative incident with known attributes, such as, duration and severity (or capacity reduction) are considered for each link. It is also assumed that incident occurrence rate for each link in the network is known. Note that

[2] Abbas, M., and McCoy, P. 1999. Optimizing VMS locations on freeways using genetic algorithms. Private communication presented at the 78th Annual Meeting of Transportation Research Board, National Research Council (US), Washington, D.C.

[3] Chiu, Y., Huynh, N., and Mahmassani, H. 2001. Determining optimal locations for VMS's under stochastic incident scenarios. Private communication presented at the 80th Annual Meeting of Transportation Research Board, National Research Council (US), Washington, D.C.

both incident attributes and occurrence rate can be estimated using historical incident data. As a result, the expected number of incidents that could occur on a specific link a over a specific period p, n_a^p can be calculated as follows:

$$[1] \qquad n_a^p = \Delta^p x^p l_a r_a$$

where Δ^p is the duration of time period p, which is assumed to be known; x^p is the original link arrival rate (vehicles/h); l_a is the length of link a (km); r_a is the incident rate for link a (number of incidents per million vehicle kilometres).

- Drivers' routing decisions under an incident can be captured by a diversion model. A detailed discussion on this assumption will be provided later.
- No other traffic information sources are available to the drivers.

With these assumptions, we can now formally define the CMS location problem as follows: identify the location of a given number of CMS in a road network so that the following objective function for travel time savings for all p time periods and all links is maximized:

$$[2] \qquad \mathrm{TTS}(z) = \sum_p \sum_a \mathrm{TTS}_a^p(z)$$

where z is a vector representing CMS location solution, $z = \{z_a\}$, where $z_a = 1$ if link a is allocated with a CMS, 0 otherwise. Note that Σz_a is the number of CMSs to be allocated and $\mathrm{TTS}_a^p(z)$ is the expected travel time savings for a specific link a during time period p, which is a function of the location solution, f. Equation [3] can be used to estimate $\mathrm{TTS}_a^p(z)$ as follows:

$$[3] \qquad \mathrm{TTS}_a^p(z) = (D_a^p - \hat{D}_a^p) n_a^p$$

where D_a^p is the total vehicle delay caused by a given incident on link a, during time period p, without CMS information (vehicles/h). Note that D_a^p is a function of location solution (z) and $\hat{D}_a^p$ is the total vehicle delay caused by a given incident on link a, during time period p, with CMS information (vehicles/h). Note that $\hat{D}_a^p$ is a function of location solution (z). The expected number of incidents on link a during time period p (eq. [1]) is represented by n_a^p.

Incident delay without information on changeable message signs

The first element in the proposed formulation (eqs. [2] and [3]) is the total delay that drivers would experience when traversing a specific link (a) in a specific period (p) if an incident had occurred, but the drivers were not informed, D_a^p. This total user delay can be estimated using a time-dependent deterministic queuing model (HCM 2000). Depending on the incident occurrence time and the level of traffic demand at the incident period(s), different queue formation patterns could develop, which can be classified into three cases: case I, off-peak; case II, peak under-capacity; or case III, peak over-capacity. The queuing diagrams for these cases under specific conditions are shown in Fig. 1, where the shaded area represents the total user delay (D_a^p).

The incident occurrence time, denoted by t^o (h), is defined for each of the three cases as follows:

$$[4] \qquad t^o = \begin{cases} 0 & \text{case I} \\ [0 \to t^p] & \text{case II, case III} \end{cases}$$

where t^p represents the end of the peak period p under consideration and start of the off-peak period p + 1. Note that for the case I (off-peak), it is assumed that all links are under-saturated in normal traffic conditions and the incident occurrence time has negligible effect on the user delay. Therefore, to simplify calculations, only incidents that occur at the beginning of the time period will be considered ($t^o = 0$). It is further assumed that, for the peak period cases (case II and III), the incident occurrence time is uniformly distributed from zero to the period duration under consideration. To determine a value for user delay for the peak period cases, a range of incident occurrence times must be considered and the user delay results must be averaged for each of these occurrence times. An arithmetic mean of user delays for a given number of incident occurrence times is evaluated. This same method is applied in the following section for estimating delay with information.

The cumulative arrivals and departures at the incident occurrence time (t^o), denoted by N_{arr}^o and N_{dep}^o, respectively, can be calculated using eqs. [5] and [6], as

$$[5] \qquad N_{\mathrm{arr}}^o = x^p t^o \qquad \text{case I, case II, case III}$$

$$[6] \qquad N_{\mathrm{dep}}^o = \begin{cases} x^p t^o & \text{case I, case II} \\ S t^o & \text{case III} \end{cases}$$

where x^p is the flow rate on specific link during normal traffic conditions (nonincident) during period p. The arrival (flow) rate is assumed to be known for each of the time periods (vehicles/h); and S is the maximum capacity of the specific link (vehicles/h).

Note from eq. [5] that, for the under-capacity cases (case I and II), the cumulative arrivals are the same as the cumulative departures at the incident occurrence time. For case III, however, the cumulative number of arrivals is greater than the cumulative number of departures because the link is over capacitated.

Similar to the occurrence time, the time of incident clearance (t^c) and the corresponding cumulative number of departures (N_{dep}^o) is defined for each of the three cases as follows:

$$[7] \qquad t^c = t^o + \tau \qquad \text{case I, case II, case III}$$

$$[8] \qquad N_{\mathrm{dep}}^c = N_{\mathrm{dep}}^o + C\tau \qquad \text{case I, case II, case III}$$

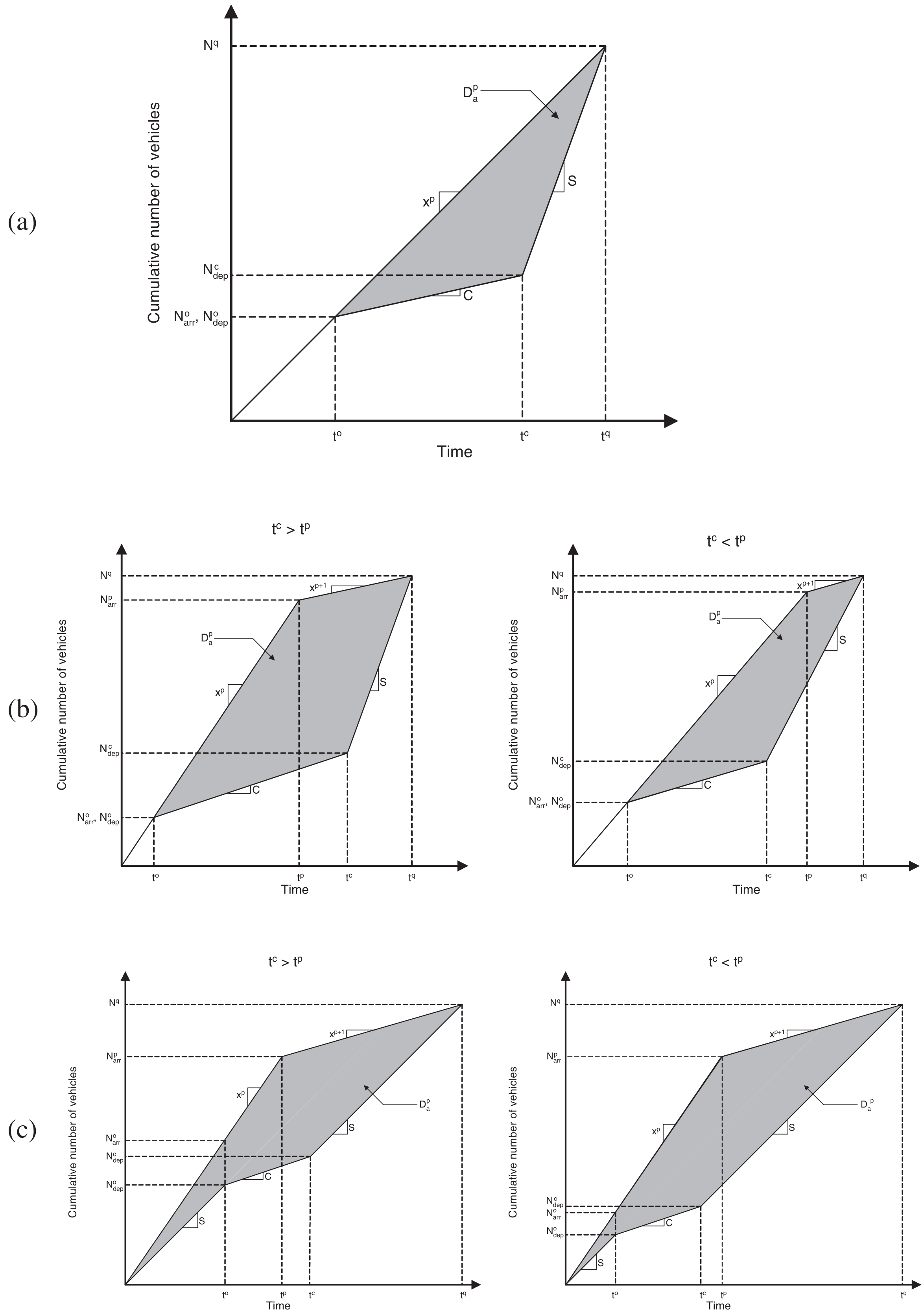
where τ is the time required to clear the incident (h); C is the reduced capacity of the link during the incident (vehicles/h). Note that case-specific values could be used for C.

Based on the assumption that the original arrival rate for a specific link (x^p) and the duration of each time period (t^p) are both known, the cumulative number of arrivals at the end of the peak period p under consideration, denoted by N_{arr}^p can be determined from the following equation:

$$[9] \qquad N_{\mathrm{arr}}^p = x^p t^p \qquad \text{case II, case III}$$

Recall that for the off-peak incident case, case I, the incident is assumed to have occurred at the start of the time period under consideration. Therefore, for case I, the end of the

Fig. 1. Vehicle queuing diagram under incident conditions for (*a*) case I, (*b*) case II, and (*c*) case III.



time period will occur after the incident queue clearance time. If the time required to clear the incident queue is long and (or) the accident occurrence time is close to the end of the period, the queue will extend to the next period. As a result, the uniform arrival assumed in case I is no longer valid and case II should be used.

The incident-queue-clearance time (t^{q}) and the cumulative number of vehicles at the queue clearance time (N^{q}) are determined from the following equations:

$$[10]\qquad t^{q}=\begin{cases}\dfrac{(S-C)t^{\mathrm{c}}}{(S-x^{p})} & \text{case I}\\ \dfrac{x^{p}(t^{p}-t^{\mathrm{o}})-x^{p+1}t^{p}+(S-C)t^{\mathrm{c}}+Ct^{\mathrm{o}}}{(S-x^{p+1})} & \text{case II}\\ \dfrac{(x^{p}-x^{p+1})t^{p}+(S-C)(t^{\mathrm{c}}-t^{\mathrm{o}})}{(S-x^{p+1})} & \text{case III}\end{cases}$$

$$[11]\qquad N^{q}=\begin{cases}x^{p}t^{\mathrm{q}} & \text{case I}\\ x^{p}t^{p}+x^{p+1}(t^{\mathrm{q}}-t^{p}) & \text{case II, case III}\end{cases}$$

Note that, with eqs. [4]–[11], all the points of the queuing diagram can be determined for each of the three cases.

It is important to note that the delay estimation methodology discussed above does not account for queue spill-back and its possible effects on delay estimation. Queue spill-back may cause three possible effects on delay estimation. The first effect is that queue spill-back may block neighbouring intersections, which would then reduce the capacity of neighbouring links and cause additional delays. Queue spill-back may also induce traffic diversion from the incident link, which would lead to an arrival rate at the incident link lower than what would normally be expected under non-incident conditions. Lastly, queue spill-back may block diversion access points, such as, freeway off ramps, which will prevent drivers from alternating their routes. These effects are difficult to accurately represent in a delay estimation model, thus, requiring detailed knowledge of network geometry and driver behaviour.

Incident delay with changeable message sign information

The second component in eq. [3] is the total delay, $\hat{D}_a^p$, caused by an incident on a specific link, a, in a specific time period, p, provided the drivers had been informed about the incident by a given set of CMSs located over the network. As discussed previously, this total delay depends on how much traffic will divert to alternative routes, which is a function of many factors, such as, CMS location, incident characteristics, traffic pattern, and availability of diverting points and alternative routes. In this research, we propose an iterative method that takes into account the time-dependent nature of incident delay and traffic routing.

The proposed method can be best illustrated by using a simple example. Consider that a total of three CMSs will be located on a network shown in Fig. 2*a*. For any given CMS location solution (e.g., the one shown in the figure), we need to determine the total incident delay if an incident occurs on a specific link. Figure 2*b* shows the corresponding queuing diagram for the traffic of the incident link, which is based on the simplest off-peak situation (case I) described in the previous section. An incident occurs at time t^{o}, and after a time lag, which includes incident detection time, information processing time, and CMS activating time, diversion because of CMS starts at time t^{s}. Note, however, that the incident link does not experience a reduction in arrival rate until $t^{s}+T_1$ because of the time lag from the closest CMS (CMS 1) to the incident link. The same is true for additional reduction in arrivals caused by CMS 2 and CMS 3.

Determining the actual reduced flow rate is not straightforward. This is because the reduction in flow rate depends on the proportion of the traffic, or more accurately, the path flow traversing the CMS and the incident link, diverted because of CMS information; whereas traffic diversion is a function of travel time savings, which in turn depends on how much traffic is diverted. We solve this interdependency problem by constructing the reduced flow rate curve from left to right starting at $t^{s}+T_1$ at a small time interval (e.g., 5 min). At the start of each interval, the proportion of flow diverted for each O–D pair is determined using the diversion model (discussed in the following section) on the basis of the expected delay for a vehicle arriving at the time instance. The cumulative number of arrivals is then calculated based on the reduced arrival rates of all O–D flows passing the CMS and the incident link. This cumulative arrival function is then used to estimate the expected delay for the subsequent interval. This process continues until the reduced arrival curve, $\hat{x}(t)$, intersects with the cumulative departure curve.

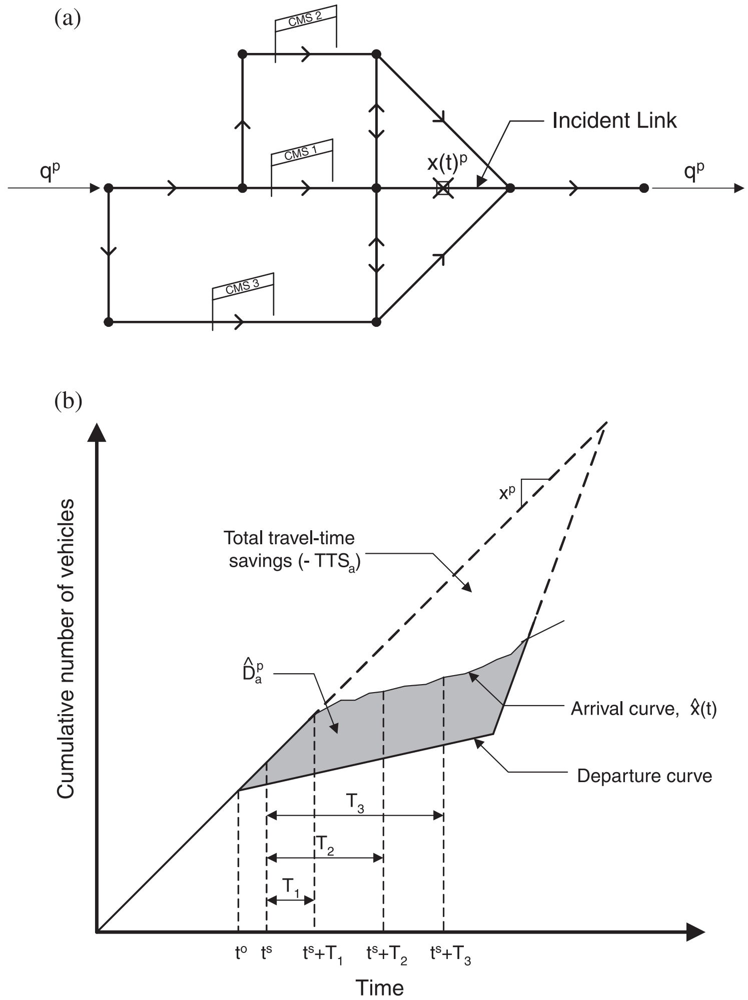
Diversion model

As described in the previous section, to quantify the incident delay under CMS information, a diversion model is required to predict the number of vehicles that would divert to alternative routes owing to message activation of the CMS during incident conditions. When a driver is provided with information from a CMS that an incident has occurred along the intended travel path, he (she) makes a decision to either stay on the same route or divert to an alternative route. This decision depends on various factors, such as, severity of the incident, current extent of queue caused by the incident, the driver's experience and familiarity of the network, and incident characteristics delivered via the CMS. Therefore, modeling the underlying decisions is a significant challenge because of the behavioural complexity of the drivers' response to incidents and incident information (Wardman et al. 1997). In the two existing studies on the CMS location problem, Abbas and McCoy[2] assumed a constant fixed diversion rate regardless of the availability of alternative routes, severity of incidents, and various other factors; Chiu et al.[3] applied a bounded route choice model, assuming a driver would divert to an alternative route if the expected travel time to be saved exceeds a certain threshold.

In this study, we propose to use a simplified-discrete-choice model to capture the major characteristics of drivers' response behaviour under incident conditions. The model was motivated by the empirical work of Wardman et al. (1997), and it was assumed that the probability for a driver to choose to divert depends on the expected travel time saving from diverting with the following logit form:

$$[12]\qquad P_{k,m}(t)=\frac{1}{1+\mathrm{e}^{\alpha-\beta S_{k,m(t)}}}$$

Fig. 2. Incident delay under changeable message signs (CMSs) information. (*a*) Location of three CMSs on a network and (*b*) corresponding queuing diagram for the traffic of the incident link.



where $P_{k,m}(t)$ is the probability for a vehicle, arriving at time t and traveling through CMS k on path m, to divert to an alternative route; $S_{k,m}(t)$ is the travel time savings ratio defined in eq. [13] as the expected savings of using the alternative route divided by the travel time of the alternative route; and α and β are model parameters.

The travel time savings ratio, $S_{k,m}(t)$, based on the expected delay that vehicles joining the incident queue will experience and the travel time through the shortest alternative route, is defined as follows:

$$[13] \qquad S_{k,m}(t) = \frac{T_{k,m}(t) - T^*_{k,m}}{T^*_{k,m}}$$

where $T_{k,m}(t)$ is the expected travel time a vehicle joining the queue will experience at time t (hours); $T^*_{k,m}$ is the travel time (h) of the shortest alternative route, not traversing the incident link, from CMS k to the destination node of path m.

The expected travel time experienced by a vehicle joining the queue, $T_{k,m}(t)$, is based on two components. The first component is the expected travel time from the CMS link, through the incident link to the destination node of the path under incident-free conditions. The second component is based on the expected queuing delay for a vehicle if it were to continue on its original path.

The diversion model represented by eq. [12] suggests that the proportion of vehicles that would divert increases as the travel-time savings increase. The relationship between diversion probability or rate and travel time savings is intuitively correct: the higher the travel time savings, the higher the probability for a vehicle to divert; drivers are usually reluctant to change routes with a small percentage of savings. For example, for the case of $\alpha = 5$ and $\beta = 5$, when an alternative route is estimated to provide 50% travel-time savings, there would be 50% chance for the driver to make a diversion. The parameters α and β are essentially to model variations in drivers' characteristics (e.g., aggressiveness) and information attributes (e.g., types, level of reliability, frequency, etc.). Realistic estimates of these model parameters could be obtained through a statistical analysis of survey results, as in Wardman et al. (1997). A sensitivity analysis is performed in this study to evaluate the potential impact of these parameters on the final CMS location solutions.

Arrival time and reduced flow rate because of a CMS may be determined by applying eq. [12], derived for a single path, to all paths, as follows:

$$[14] \quad \hat{x}_k(t) = \sum_m f_{k,m}[1 - P_{k,m}(t)]$$

where $x_k(t)$ is the flow on the incident link, or reduced arrival rate, for vehicles that traverse both CMS k and the incident link (vehicles/h); and $f_{k,m}$ is the flow on the mth path passing both CMS k and the incident link (vehicles/h).

In determining the reduced arrival rate at a given link because of all CMSs, it is assumed that drivers would defer their decision until they reach the CMS that is closest to the incident link. That is, for a path flow that traverses several CMSs, only the CMS closest to the incident link has an effect on the diversion rate. An activating zone is also considered so that only CMS within a certain distance of the incident will display information.

Path-based traffic assignment

To estimate the delay caused by incidents, network traffic for the time periods of interest must be obtained first. Typically, network traffic is obtained using a link-based traffic assignment method (e.g., Frank–Wolfe method), which produces link flow estimates by assigning a matrix of origin–destination (O–D) flows to individual links based on Wardrop's user-equilibrium (UE) assumption (Sheffi 1985). The resulting information (i.e., link traffic flow) is, however, not sufficient for the proposed benefit model, as it requires not only the traffic flow on individual links but also the individual path flows between the origins and destinations. Path flows are needed in both the prediction of traffic diversion and the alternative route travel times, as discussed in the previous sections.

Generally, there are two classes of path-based assignment methods: incremental and user equilibrium. The incremental method performs successive all-or-nothing assignments on a parsed O–D trip matrix (Sheffi 1985). As the incremental assignment does not always result in a UE condition, therefore, it will not be considered further. Several user-equilibrium path-based methods for traffic assignment have been developed, including a modified Frank–Wolfe approach by Chen and Lee[4]. However, the gradient-projection (GP) method proposed by Jayakrishnan et al. (1994) for traffic assignment has so far proven to be the most efficient one and will be used here. A detailed discussion can be found in Henderson (2004).

Effect of diversion on alternative routes

The presented diversion model approximates the rate at which drivers divert from their originally intended route to one or more alternatives during incident conditions. However, no prediction is made as to what those alternative routes are, since there are presently no reliable route choice models during traffic equilibrium disruption. Additionally, the increased travel time experienced by drivers on these alternatives are not considered in the travel time savings calculation of eq. [2].

To partially account for this impact, an alteration to the diversion rate and incident link-based travel-time savings is proposed that reassigns diverted traffic to the shortest alternative route between the CMS link and the trip destination. A diversion equilibrium is approximated through successive iterations of $T_{k,m}(t)$ and $T^*_{k,m}$ in eq. [13]. During the iterations, link-travel times are updated to reflect volume changes resulting from traffic diversion. These updated link-travel times may then be included in eq. [2] to adequately reflect the negative impact of diverted traffic on otherwise unaffected links, that is,

$$[15] \quad \mathrm{TTS}_{\max} = \sum_p \sum_a (\mathrm{TTS}_a^p - \mathrm{TTI}_a^p)$$

where TTI_a^p is the travel time increase on alternate routes caused by incident on link a during time period p.

Solution procedure

The CMS location optimization problem formulated in the previous section belongs to the general class of location problems that have been researched extensively by the operations research (OR) community. The general location problem is computationally intractable and is commonly solved using heuristic algorithms. In this paper, we applied a greedy algorithm that allocates one CMS at a time and fix the allocated CMS in the subsequent steps of allocating other CMSs. A detailed description of this algorithm is available in Henderson (2004), which also includes a more advanced solution method based on genetic algorithm (GA). The greedy method was chosen because it most closely resembles the administrative decision-making process that incrementally adds changeable message signs to a road network. The incremental method was used for traffic assignment to identify path flows. This method was observed to produce link volumes close to that of the Frank–Wolfe equilibrium traffic assignment and a path-based traffic assignment procedure for the sample network used in the following section. The overall solution process has been implemented in a software tool called OptimalCMS and used in the following sensitivity analysis.

Sensitivity analysis

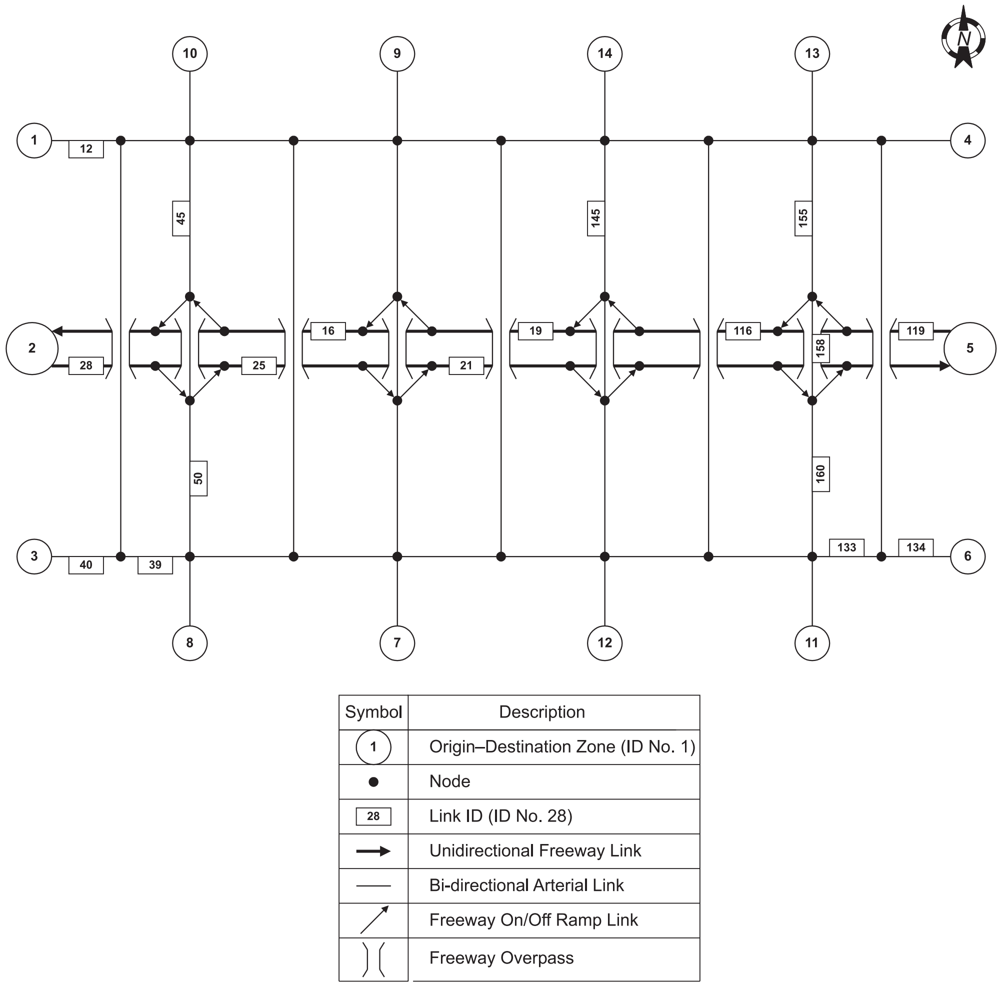
The proposed model for evaluating the expected reduction in delay, due to the installation of CMSs, is deterministic in nature, assuming perfect information on input parameters, such as, O–D demand, incident conditions, and traffic diversion behaviour. In practice, however, most of these parameters are inherently uncertain because of various factors, such as, insufficient relevant data to estimate the model parameters, errors in both the raw data and model specification, and errors in predicting future traffic demands and conditions. The objective of this section is to quantify the potential effects of these variations on CMS location. These effects will be explored using both a hypothetical network (case A) and a section of Highway 401 in Toronto (case B).

Case A — sample network

Figure 3 shows the layout of one of the road networks used in our sensitivity analysis, which is a modified version

[4] Chen, A., and Lee, D. 1999. Path-based algorithms for large scale traffic equilibrium problems: a comparison between DSD and GP. Private communication presented at the 78th Annual Meeting of Transportation Research Board, National Research Council (US), Washington, D.C.

Fig. 3. Sample network.



of a sample network used in the integration simulation model (M. Van Aerde 2000). The network represents a freeway-arterial system consisting of 56 nodes and 124 links. An east–west freeway is located in the middle of the network surrounded by arterial links with a total length of 6.4 km.

Also shown in the network are 14 O–D zones as trip generators. Four time periods are considered: AM peak, midday off-peak, PM peak, and overnight off-peak with period durations of 2, 7, 3, and 12 h, respectively. The O–D demand pattern is created so that a large number of trips travel eastbound (from zone 2 to zones 4, 5, and 6) during the AM peak period and, conversely, a large number of trips travel westbound (from zone 5 to zones 1, 2, and 3) during the PM peak period. The rate of AM demand is monotonically higher than the PM demand. Also, the demand matrix for modelling uncertainty in traffic demand and driver behaviour is different (larger demand) than the demand matrix for modelling uncertainty in incident attributes.

The base incident rate was set at 2.9 incidents per million vehicle kilometres according to a report from National Highway Traffic Safety Administration (TRB 2000). Equation [1] was used to convert this value, r_a, to incidents over a given time period, n_a^p. Other base values used for this sample network and applied to all incident links are capacity reduction of 0.8, incident duration of 30 min, detection time of 10 min, and processing and CMS activating time of 5 min. These values are similar to incident characteristics used in other relevant literature (Abbas and McCoy[2]; Chiu et al.[3]). Impact of variations in these parameters on optimal CMSs locations is also analysed in this section.

Comparison to heuristic methodology

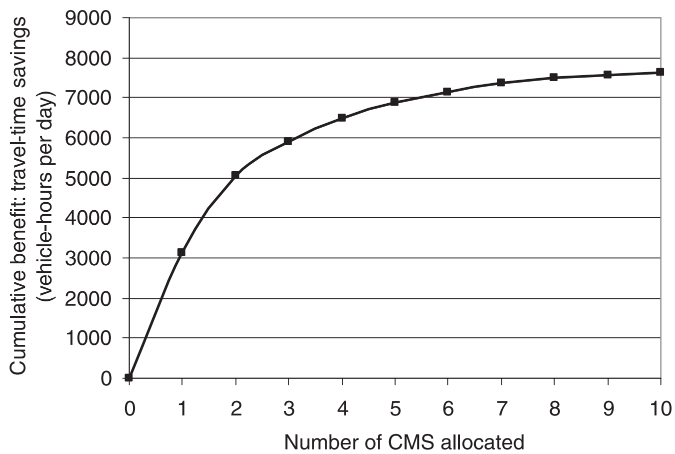
In the absence of a methodical approach for optimizing CMSs locations, traffic managers usually locate CMSs in a heuristic ad-hoc way based on freeway link traffic volume and diversion opportunities at downstream off-ramps. To illustrate the difference between the proposed model and this ad-hoc approach, we consider the greedy allocation results for the base case of the sample network. Table 1 shows the order of CMS allocation based on the proposed model and the revised order based on link traffic volume. It is reasonable to assume that there is a good opportunity for traffic to divert to an alternative route from each of these links, since they were all selected in the greedy allocation process.

As seen in Table 1, there is a significant difference between CMS allocation based on traffic volume and allocation based on the proposed model. Using the proposed model, freeway links with the highest volume are not necessarily first to be assigned a CMS.

Table 1. Changable message signs (CMSs) allocation order, proposed model versus link volume.

Allocated CMSs at greedy iteration No.	Link ID (base case)	Daily traffic volume	CMSs order based on daily traffic volume
1	119	22 500	6
2	28	21 295	7
3	116	53 310	1
4	25	47 930	2
5	155	20 345	8
6	45	17 980	9
7	50	14 350	10
8	158	24 105	5
9	16	47 790	3
10	21	46 750	4

Fig. 4. Marginal benefit of changeable message signs (CMSs) allocation.



Optimal number of changeable message signs

Figure 4 shows the relationship between the total benefits of CMSs and the total number of CMSs installed in the network. A significant increase in the total network benefit resulted from the addition of the first 3 CMSs to the network (an increased travel time saving of about 62% from 1 CMS to 2 CMSs). The marginal network benefits from adding more CMSs tend to become less significant and level off after the third CMS is added. This result is expected because when the coverage of CMSs reaches to a certain level, majority of the O–D flows will be covered by the CMSs, thus the benefits of adding more CMSs become less. This general phenomenon of diminishing returns is commonly seen in economics and suggests the potential of obtaining the optimal number of CMSs for a given network is by trading off the installation costs and the resulting benefits.

Consideration of multiple time periods

Most existing models for optimizing CMS locations consider only a single time period, AM or PM peak, and ignore the traffic exposure for the remainder of the day. The proposed model, however, is able to consider the entire day in the optimization process. The CMS allocation results based on a single time period compared with the results for all time periods are shown in Table 2.

It is evident that only considering a single time period in the optimization process does not produce the best CMS locations, since demand is directional and the highest volume freeway links during the AM peak period are not the highest volume freeway links during the PM peak period. Also, incidents during the off-peak period had a negligible impact on the final location solutions of the optimization process, as the estimated benefit was several orders of magnitude smaller than the benefit predicted during the peak periods.

Uncertainty in traffic demand

The uncertainty in traffic demand was modelled by considering random fluctuations in the O–D matrix. Variations of +10%, –10%, ±5%, ±10%, and ±20% to the base O–D demand matrix were evaluated. The variation in each case was developed by increasing or decreasing each entry in the base O–D matrix (AM, PM, and off-peak) by a certain amount. For example, for the case of +10% variation, each demand entry was increased by a random amount from 0% to 10% inclusive. The –10% variation in demand was determined in a similar manner. The ±5%, ±10%, and ±20% variations in demand were also calculated similar to the +10% variation with the exception of the interval of equal likelihood, which is increased from 0% → 10% to –10% → 10% for the ±10% variation. The results of the greedy allocation procedure for each of the demand variations are shown in Table 3.

As seen in Table 3, variations in demand do not significantly effect the allocation of changeable message sign locations. A slight variation in CMS allocation order was observed, but this was only for a few locations with marginal travel time savings benefit.

Uncertainty in incident conditions

Four incident attributes, including incident rate, incident duration, incident occurrence time, and capacity reduction, were modelled for uncertainty. Since the incident occurrence time, as mentioned earlier, is assumed to be distributed uniformly over the time period of concern, it is not considered hereafter. The uncertainty in incident rate was modelled by considering random fluctuations in the link exposure-based incident rate. Fluctuations of +10%, –10%, +50%, –50%, +100%, and –100% were considered and developed, as previously described for other variations. The CMS location results for variations in link incident rates are shown in Table 4.

The optimal CMS locations are insensitive to smaller variations of 10% and 50% in the link incident rate, however, larger variations seem to affect the allocated CMS locations. The reason for the effect on CMS locations is the linear relationship between incident rates and travel-time savings; the increase in travel-time savings is directly proportional to the increase in link-incident rates. The link exposure-based incident-rate variable may be stochastic instead of deterministic as was assumed.

An 80% reduction in link capacity was used for the base case. Four additional cases were generated: the first two cases involved monotonic reductions of 40% and 60% for all links; the other two cases had variations similar to the random fluctuations in congestion level. A random decrease in capacity with equal likelihood from 50% to 70% inclusive

Table 2. Greedy allocation results based on time period.

	All time periods		AM peak		Midday		PM peak		Overnight	
Greedy iteration No.	Link	MS	Link	MS	Link	MS	Link	MS	Link	MS
1	119	3123	28	1941	45	9	119	3120	16	0
2	28	1942	25	556	16	2	116	841	45	0
3	116	849	45	223	155	1	133	407	125	0
4	25	560	39	142	50	1	155	266	155	0
5	155	406	21	55	160	1	19	63	50	0
6	45	268	145	30	21	0	176	28	160	0
7	50	224	75	29	116	0	55	20	105	0
8	158	142	125	9	125	0	58	14	174	0
9	16	64	12	7	25	0	16	9	68	0
10	21	55	35	4	106	0	145	5	80	0

Note: MS, marginal savings (vehicle·h/d).

Table 3. Changeable message signs (CMSs) allocation results for variable congestion levels.

	Base case		+10%		–10%		–5% ~ +5%		–10% ~ +10%		–20% ~ +20%	
Greedy iteration No.	Link	MS	Link	MS	Link	MS	Link	MS	Link	MS	Link	MS
1	119	3123	119	3489	119	2997	119	3254	119	3628	119	2448
2	28	1942	28	2087	28	1675	28	1921	28	1918	28	2123
3	116	849	116	933	116	812	116	865	116	891	116	772
4	25	560	25	601	25	508	25	562	25	563	25	575
5	133	406	133	449	133	372	133	422	133	415	133	331
6	155	268	155	308	155	242	155	279	155	312	**45**	**248**
7	45	224	45	240	45	197	45	221	45	223	**155**	**222**
8	39	142	**50**	**144**	**50**	**134**	39	147	39	144	**50**	**129**
9	19	64	**21**	**72**	19	57	19	58	19	63	**21**	**70**
10	21	55	**19**	**66**	21	51	21	55	21	55	**19**	**63**

Note: MS, marginal savings (vehicle·h/d). Values in bold represent changed links for CMSs.

for one case and 40% to 80% for the other case was generated independently for each of the links. The CMS location results for capacity reduction variations were identical for the first six CMSs allocated and only the order of allocation for the last four varied from the base case. This indicates that the optimal CMS location is insensitive to uniform changes in capacity reduction and relatively insensitive to random fluctuations in capacity reduction. Again, a slight variation in CMS allocation order was observed but only for a few locations that produced a small marginal benefit.

Variations in incident duration were evaluated using 30 min for the base incident duration case and uniform durations of 40 and 50 min for additional uniform-incident duration cases. Random fluctuations, similar to those discussed for capacity reduction, were modelled using two additional random duration cases with equal likelihood for durations between 30 and 50 min inclusive for the first random case and between 20 and 60 min inclusive for the second random case. The results of the allocation procedure, similar to those of variations in link capacity reduction, show identical CMS locations for the first six CMSs allocated and differing order for the last four CMSs allocated. Therefore, with the exception of a few locations with marginal benefit, the allocated CMS location is insensitive to both monotonic and random variations in incident duration.

Case B — Toronto network

The Toronto case will now be considered for additional analysis. The complete dataset (Fig. 5) provided by the Ministry of Transportation of Ontario consists of 14 160 nodes, 37 386 links, and 69 448 O–D pairs extracted from the EMME/2 transportation planning software. This study area is prohibitively large for model execution, therefore, a smaller network (Fig. 5) mainly comprised of Highway 401 and its nearby arterials were extracted for computational analysis. This reduced the number of nodes, links, and O–D pairs to 961, 2363, and 6149, respectively, with the longest of these links being 3.45 km and the shortest being 0.04 km. Also, the link free-flow speeds range from 40 ~ 70 km/h for the arterials and ramps and 110 km/h for the freeway links. The smaller study area is approximately 23.2 km in the east–west direction, from west of Highway 410 to east of Allen Road, and 17.4 km from the most southerly node to the most northerly node.

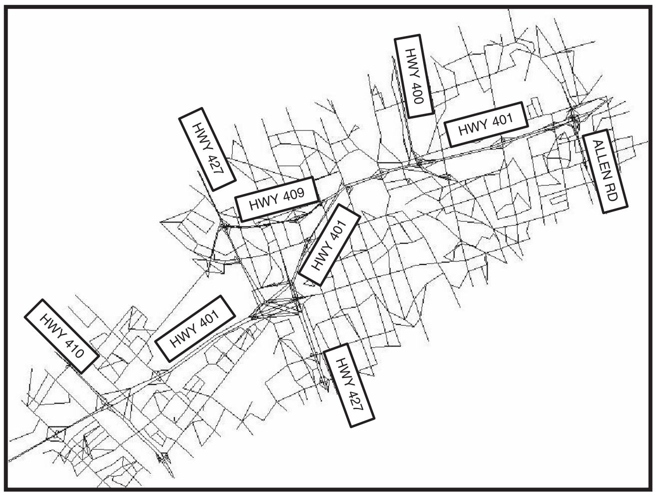
The Highway 401 demand matrix was determined by performing path-based assignment for the complete dataset and saving the path information for each O–D pair. The paths, referenced as a succession of nodes, were truncated at the edges of the study area. The truncation was performed by eliminating all nodes up to the first node that was part of the smaller network. The demand for this new O–D pair was

Table 4. Changeable message signs (CMSs) allocation results for variable link incident rates.

	Base case		+10%		+50%		–50%		+100%		–100%	
Greedy iteration No.	Link	MS	Link	MS	Link	MS	Link	MS	Link	MS	Link	MS
1	119	559	119	545	119	735	119	549	119	605	119	283
2	28	417	28	444	28	375	28	381	28	461	28	218
3	116	138	116	134	116	173	116	114	116	160	**155**	**56**
4	25	103	25	107	25	94	25	83	25	116	**45**	**41**
5	155	49	155	48	155	65	155	62	**45**	**50**	**116**	**33**
6	45	43	45	46	45	39	45	44	**155**	**44**	**50**	**29**
7	50	25	50	27	**158**	**33**	**158**	**31**	50	26	**158**	**19**
8	158	24	158	24	**50**	**22**	**50**	**26**	**160**	**22**	**21**	**17**
9	16	15	**21**	**15**	**19**	**18**	16	13	16	22	**25**	**4**
10	21	15	**16**	**15**	21	16	21	12	21	17	**19**	**3**

Note: MS, marginal savings (vehicle·h/d). Values in bold represent changed links for CMSs.

Fig. 5. Highway 401 network extracted from the City of Toronto network.



then set to the associated path flow. Demands with common origins and destinations were combined to reduce the size of the derived demand matrix. The peak hour demand (AM) for this matrix is approximately 152 000 vehicles and the associated daily demand is approximately 948 500 vehicles.

Optimal changeable message signs locations and marginal benefits

Similar to the previous analysis, we first run our location optimization program to locate ten CMSs to this network. It was found that most of the CMS have been allocated to freeway links upstream of an interchange. This is expected as interchanges represent an excellent diversion opportunity. The first five CMS locations are described as follows (Fig. 5).

- First CMS is located on westbound Highway 401 just before Allen Road. This location captures most of the westbound freeway traffic while providing a diversion opportunity to Allen Road.
- Second CMS is located on westbound Highway 401 just after the on-ramp from southbound Highway 400. Highway 400 as well as Highway 401 traffic pass this point and can be informed of conditions on the Highway 401 main route and Highways 409 and 427 alternative route.
- Third CMS is located on the eastbound Highway 401 after the Highway 427 interchange. The Highways 401 and 427 traffic may divert at a minor interchange downstream of the CMS link.
- Fourth CMS is located on eastbound Highway 401 at the Highway 400 interchange, which has a high traffic volume but poorer diversion opportunities than the first three CMS allocated.
- Fifth CMS is located at the beginning of the eastbound Highway 409, providing two excellent diversion opportunities through eastbound Highway 409 or southbound Highway 427. However, the traffic volume is much lower than the first four locations that CMS were located.

Similar to the results shown in Fig. 4, our analysis of this realistic network also shows that the total benefits of CMS is an increasing function of the total number of CMS installed in the network. The general trend is the same with a very high benefit for the first few CMS installed, followed by decreasingly marginal gains. Practically no further network benefit is observed between the ninth and tenth CMS installation. The benefit of adding CMS to Highway 401 is approximately 1.52 million vehicle·h/year when considering an α and a β of 5. This benefit is the difference between the incident-induced delay without CMS information (4.44 million vehicle·h/year) and the incident-induced delay in the presence of CMS information (2.92 million vehicle·h/year).

The network benefit may be converted to a dollar value by assuming a value of time, e.g., CAN\$10/h. Using this value of time, the benefit attained by installing one CMS, approximately 640 000 $\rightarrow$ 1 010 000 vehicle·h, may be converted to a range of CAN\$6.4 – CAN\$10.1 million. After the installation of the second CMS the benefit increased to the range of CAN\$8.5 – CAN\$11.9 million. Again, a cost-benefit analysis would be complementary in determining the optimal number of CMS to install.

Uncertainty in diversion model

The diversion model contains a high degree of uncertainty. To identify the impact of the diversion model on CMS allocation, CMSs were allocated on the basis of different α and β values. Recall from eq. [12] that increasing α will monotonically decrease the diversion rate for all paths and increasing β will increase the diversion rate, depending on the travel time for the shortest alternative route. Therefore, a comparison of the marginal benefits of each alloca-

Table 5. Links allocated with changeable message signs (CMSs) under variable diversion model parameters.

Greedy iteration No.	Model parameters $\alpha = 5; \beta = 5$	$\alpha = 5; \beta = 10$	$\alpha = 10; \beta = 5$	$\alpha = 10; \beta = 10$
1	7628	7628	7628	7628
2	9980	7547	**29007**	9980
3	7547	9980	**7593**	7547
4	9986	9986	**7607**	9986
5	29486	**29469**	**9969**	29486
6	28974	28974	7615	28974
7	7615	10250	10250	**9774**
8	9972	**7559**	**10089**	**10421**
9	10250	7615	**28039**	**7575**
10	7597	7597	**25008**	**9966**

Note: Values in bold represent changed links compared to the first solution ($\alpha = \beta = 5$).

tion procedure is not relevant. Instead the focus will be on a comparison of the chosen locations (Table 5).

As seen in Table 5, the CMS locations are highly sensitive to changes in the diversion model parameters. Generally, as β increases the actual travel time savings ratio, $S_{k,m}$, is less significant because $\alpha - \beta S_{k,m}$ becomes very large in the negative direction and $e^{\alpha-\beta\, Sk,m}$ approaches zero, so more and more incident links have close to 100% diversion. Since the savings ratio becomes less significant, the volume and number of incident links for which the CMS is effective becomes the predominant factor in locating the CMS. Therefore, as β increases the CMSs are located closer to the start of the paths to maximize the number of links and traffic volume for which they are effective. Also, it was observed that these CMSs links have longer alternative routes and lower travel-time savings ratios compared with CMSs links allocated at lower values of β.

Inclusion of alternative path travel time

As previously indicated, the increased delay incurred on alternative routes may be a factor in CMS-location decisions. Therefore, a comparison was made between chosen CMS locations with and without the inclusion of the negative impact that a diversion has on alternative route travel times. Recall that both methods include the travel times on alternate routes, but the consideration of alternate path travel times are used along with diverted traffic information in determining diversion rates. It was found that the inclusion of alternate path travel time slightly increases the travel-time benefit. This can be expected because the iterative approach to determining diversion rates would prevent excessive re-routing of traffic to congested alternate paths, however, the change is not enough to significantly alter the solution benefit.

The locations identified by these two approaches were also compared, and it was found that the results with and without the consideration of alternate path travel time are quite different. This result is reasonable when considering that the optimal solution may not be unique. That is, there may be many different combinations of solutions that would produce the same objective function value. Although the solutions may be different, the actual end result is the same.

Conclusions and future research

Changeable message signs (CMSs) are becoming an important component of ITS applications, such as, advanced traffic management and traveler information systems (ATMS/ATIS). By providing travelers with accurate, timely, and reliable traffic information, safety and efficiency of the road network can be improved. The effectiveness of CMSs, however, depend on how many CMSs are installed and where the CMSs are located. This paper presents an optimization model that can be used to systematically locate CMS in an integrated freeway-arterial network. The proposed model consists of four components including (1) a multi-period user-equilibrium traffic assignment procedure to estimate traffic volumes on individual links and path flows between individual O–D pairs; (2) a dynamic diversion model that relates the probability for a vehicle to divert, from an incident path to an alternative route with the potential for travel time savings; (3) a time-dependent queuing model to estimate delays with and without the presence of information; and (4) a sequential optimization model to identify the best locations for a given set of CMSs.

A sensitivity analysis on the potential impacts of the variations in various parameters on the CMSs locations has resulted in the following findings: (1) the optimal locations of CMSs are insensitive to variations or estimation errors in traffic demand and incident conditions with the possible exception of large variations in link incident rates; (2) the location algorithm is highly sensitive to the time period considered, however, off-peak periods have little effect on the optimal CMS location; (3) the location algorithm is also highly sensitive to the diversion model parameters.

The model presented in this paper is by no means complete and several modifications could be made in the future to improve the results. First, more research is needed to accurately predict diversion model parameters under incident conditions with CMSs information. Second, the benefit of other CMSs uses (e.g., environmental information) needs to be quantified before they are included in a CMS location model. Lastly, other traffic diversion models that take into account repetition effect of vehicles passing multiple CMSs should be considered.

Acknowledgements

This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

Henderson, J.M. 2004. A planning model for optimizing locations of changeable message signs. M.A.Sc. thesis, Department of Civil Engineering, University of Waterloo, Waterloo, Ont.

Jayakrishnan, R., Tsai, W., Prashker, J., and Rajadhyaksha, S. 1994. A faster path-based algorithm for traffic assignment. Transportation Research Record, **1443**: 75–83.

M. Van Aerde & Assoc. 2000. INTEGRATION: a model for simulating integrated traffic networks. Transportation Systems Research Group, Department of Civil Engineering, Queen's University, Kingston, Ont.

Sheffi, Y. 1985. Urban transportation networks: equilibrium analysis with mathematical programming methods. Prentice-Hall, Inc., Englewood Cliffs, N.J.

Shrank, D., and Lomax, T. 2002. The 2002 urban mobility report. Technical Report, Texas Transportation Institute, College Station, Tex.

TRB. 2000. Highway capacity manual (2000 update). Special Report 209, Transportation Research Board, National Research Council (US), Washington, D.C.

Wardman, M., Bonsall, P.W., and Shires, J.D. 1997. Driver response to variable message signs: a stated preference investigation. Transportation Research, Part C: Emerging Technologies, **5**(6): 389–405.