

Design and Evaluation of Operational Strategies for Deploying Emergency Response Teams: Dispatching or Patrolling

Shanjiang Zhu¹; Woon Kim²; Gang-Len Chang, M.ASCE³; and Steve Rochon⁴

Abstract: Both patrolling and prepositioned strategies for allocating emergency traffic response units have been implemented in practice. To compare the performance of both response strategies, this study has conducted an efficiency comparison based on the field data from the I-495/I-95 Capital Beltway. The extensive experimental results have revealed that the effectiveness of those response strategies varies with some critical factors, including the spatial distribution of incident frequency over different times of a day, the fleet size of the response team, the congestion level, and the available detection sources. In view of the resource constraints, the study has further presented a methodology to determine the most cost-beneficial fleet size operated with the proposed strategies, considering the marginal cost and the benefit of an additional response unit on the resulting total social benefits. The analysis results with the data from the Capital Beltway could serve as the basis for highway agencies to review and optimize their incident response and management program. DOI: [10.1061/\(ASCE\)TE.1943-5436.0000670](https://doi.org/10.1061/(ASCE)TE.1943-5436.0000670). © 2014 American Society of Civil Engineers.

Author keywords: Dispatching strategy; Patrol strategy; Incident response.

Introduction

Traffic incidents have become a major cause of congestion and a significant threat to urban mobility. The Federal Highway Administration (FHWA) (2005) found that approximately 25% of congestion in the United States is incident related. According to the latest Urban Mobility Report (Schrank et al. 2011) by the Texas Transportation Institute (TTI), traffic incidents are estimated to account for approximately 52–58% of the total delays experienced by motorists in all urban areas.

An effective incident management system not only helps to mitigate congestion through swift incident detection, response, and site clearance, but also generates significant environmental benefits by reducing fuel consumption, emissions, and potential secondary incidents. The importance of incident response management systems has been well recognized among traffic management agencies. As reported in a recent study (Schrank et al. 2011), 13 of the total 15 large metropolitan areas and 30 out of the total 32 medium urban areas list incident management systems as one key solution to mitigate traffic congestion. However, how to

maximize the incident management efficiency under available resource constraints remains a critical issue.

Most incident response strategies proposed in the literature can be classified into two categories: dispatching and patrolling systems. Some researchers (Skabardonis et al. 1998; Lou et al. 2010) investigated the freeway service patrols (FSP) program, under which incident response units will constantly roam on freeways to detect and respond to traffic incidents. The key question for such a system is how to divide a traffic network into independent patrol segments and how to assign response units to them. Some previous studies have addressed this problem either analytically (Pal and Sinha 1997) or through simulation (Pal and Sinha 2002; Ozbay and Bartin 2003). In contrast, Larson and Odoni (1981) and Pal and Bose (2009) argued that it is more efficient to deploy response units strategically and dispatch them after an incident has been detected. This system is more suitable to areas where traffic surveillance and incident detection are available, and the key question for such strategies is how to deploy and dispatch available response units so as to minimize the response time. Early studies along this direction include the p-median problem first introduced by Hakimi (1964) and the maximal coverage location problem (MCLP) proposed by Church and ReVelle (1974). Both paradigms of response strategies have their advantages in practice, depending on the incident locations and some operational constraints. The optimal design for incident response units may vary over different time periods (peak versus nonpeak) and with available resources.

Another challenge for designing and deploying incident management systems is the stochastic nature of incidents. In practice, two incidents may happen concurrently, and the response unit at the nearest depot location may not be available. Therefore, a back-up strategy must be available, and the associated cost must be considered in the system design. For example, Sherali and Subramanian (1999) introduced a new term in their objective function to represent the opportunity cost related to loss of coverage when a response team is busy. Geroliminis et al. (2009) extended the hypercube queuing model to generate optimal strategies in a

¹Assistant Professor, Dept. of Civil, Environmental, and Infrastructure Engineering, George Mason Univ., 4400 University Dr., MS 6C1, Suite 1300, Fairfax, VA 22030 (corresponding author). E-mail: szhu3@gmu.edu

²Graduate Research Assistant, Dept. of Civil and Environmental Engineering, Univ. of Maryland, 1173 Glenn Martin Hall, College Park, MD 22472. E-mail: wkim2@umd.edu

³Professor, Dept. of Civil and Environmental Engineering, Univ. of Maryland, 1173 Glenn Martin Hall, College Park, MD 22472. E-mail: gang@umd.edu

⁴Division Chief, Traffic Development and Support Division, Office of Traffic and Safety, Maryland State Highway Administration, 7491 Connelly Dr., Hanover, MD 21076. E-mail: srochon@sha.state.md.us

Note. This manuscript was submitted on December 20, 2012; approved on February 4, 2014; published online on March 31, 2014. Discussion period open until August 31, 2014; separate discussions must be submitted for individual papers. This paper is part of the *Journal of Transportation Engineering*, © ASCE, ISSN 0733-947X/04014021(12)/\$25.00.

dispatching system while considering the availability of a response unit. Another type of uncertainty is because of the variance of incident rates. Historical data of incident frequency usually exhibit a significant variance. Many existing models are built based on historical data and adopt the minimal expected response time as the design objective. Those models may perform well most of the time, but generate excessively long response times in some cases. In practice, traffic management agencies may be willing to sacrifice some savings on average response time, but to constrain the total response time in those worst cases in an acceptable range. However, such constraints are not adequately considered in existing systems.

In a previous study, Zhu et al. (2012) empirically compared the performance of the optimal incident response unit location/allocation strategy based on the p -median model with three experience-based strategies: (1) allocating available response units near high-frequency incident locations; (2) distributing them evenly over the entire coverage area; and (3) positioning standing-by units at the traffic operation center (TOC). However, all of these strategies preposition emergency response units and dispatch them if an incident is detected. Moreover, previous work by Zhu et al. (2012) did not consider the stochastic nature of incidents and the likelihood of having multiple incidents, which compromises its applicability and transferability to different regions.

Many states, such as Maryland, do not have a well-established guidance on how incident response units should be operated, and each traffic operation center decides their own rules based on engineering judgment. There have been very few empirical studies in literature on whether the patrolling or the prepositioned strategies perform better under various traffic and incident conditions. A systematic approach is needed to take advantage of the growing field data and empirically evaluate the performance of these two strategies. To bridge this gap, this paper extends this framework and compares the performance of patrolling systems with prepositioning/dispatching strategies. To ensure the effectiveness of the proposed model, the system performance during different time periods and under different resource constraints will be compared with data collected from 2006–2010 on the I-495 beltway in the state of Maryland. The potential impact of the variance of incident frequency has also been evaluated to ensure fair comparison. Findings from this research may help incident management agencies such as the Coordinated Highway Action Response Team (CHART) in Maryland to better design their response system and to adapt their response strategies within the resource constraints. The next section will present the dispatching model, followed by a discussion of data used in this study.

Performance of different incident response systems will then be discussed, along with a benefit/cost analysis.

Incident Response Models

This study considers two types of incident response systems: (1) allocating available response units strategically at several depots and dispatching them optimally if an incident has been detected; and (2) dividing the network into patrol beats and allocating response units to patrol constantly, detect, and respond to incidents.

Dispatching Model

To be consistent with previous research in the literature, this study considers the scenario of I freeway segments to be served by no more than P emergency response units (constrained by available resources), which will be located at J potential sites for response units. The variable M is the set of location sites to house the response units; thus, $M \subseteq J$. In general, the system takes the following variables as inputs: t_{ij} , travel cost from location j to segment i ; and λ_i , the average accident rate on freeway segment i during a given time period.

Both the incident and the emergency response unit should situate at one specific location on a freeway segment. To be consistent with previous literature, this study assumes that they all locate at the middle point of the corresponding segments. In practice, a segment is defined as one freeway section between two exits and with its length varying from 1.61–3.32 km (1–2 mi). Therefore, this simplification should not significantly impact the results.

This study assumes that incident frequency follows a Poisson distribution, as widely used in existing studies. Examples include those by Kweon and Kockelman (2003), Skabardonis et al. (1997), and Ma et al. (2008). Researchers (Lord and Mannering 2010) have also raised concerns on the use of a Poisson distribution, such as overdispersion and excessive zero counts, and more advanced statistical tools (e.g., zero-inflated Poisson) were proposed to address some of these concerns. However, Lord et al. (2005) argued that the fundamental incident process indeed follows Poisson trials by nature, and some of the concerns are because of “inappropriate selection of time/space scales.” Because no consensus is reached in the literature, an empirical test was conducted to evaluate the incident frequency distribution.

This study used incident data on the Capital Beltway in Maryland from 2006–2008. Fig. 1 provides an example of this incident data by comparing the empirical and theoretical incident

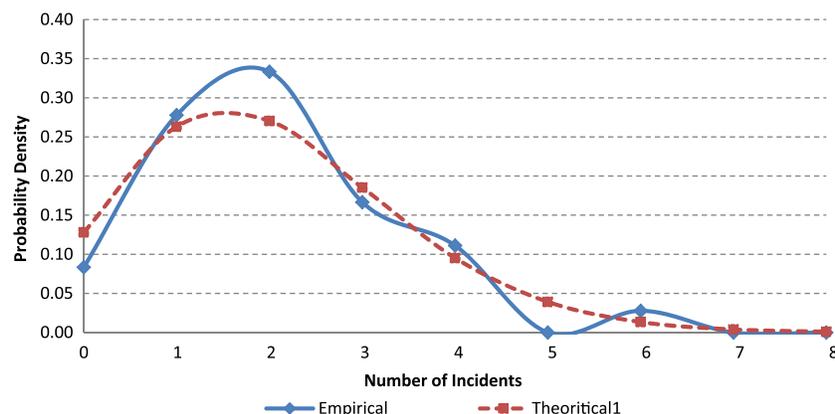


Fig. 1. Distribution of incident occurrences on exit 38 on I-495 during a.m. peak hours

Table 1. Percentage of Freeway Segments Passing the Goodness-of-Fit Test of Different Distributions on Incident Frequencies using Data Collected from the I-495 Capital Beltway in Maryland

Distributions	Time of day	
	a.m. peak (%)	Night (%)
Poisson	67	80
Normal	26.7	0
Log-normal	10	0

Note: Results are based on the 95% confidence interval.

distribution for exit 38 during a.m. peak hours. The graph suggests that the Poisson distribution well matches the field data. A more rigorous statistical test (Friendly 2000) has been conducted to test the goodness-of-fit of the Poisson distribution for each of the freeway segments during a.m. peak periods and nights. As a comparison, the same statistical tests have also been conducted for two other widely used distributions, normal distribution and log-normal distribution, using the same data set. The results are summarized in Table 1. At the 95% confidence interval, incident occurrences on approximately 67 and 80% of subsegments of the Capital Beltway follow a Poisson distribution during a.m. peak periods and at night, respectively. In contrast, incident patterns on only 26.7% of freeway segments followed the normal distribution during a.m. peak periods, and none of them followed the normal distribution at night. The results for the log-normal distribution are even worse. Therefore, the assumption of a Poisson distributed incident frequency is adopted at this stage. One may choose to use a different distribution if supported by the field data.

Given the average incident rate λ_i , the probability that an accident occurs in service segment i during a given time period is

$$\Pr(\tau_i = 1) = \lambda_i e^{-\lambda_i} \quad (1)$$

and the probability of two accidents occurring concurrently is

$$\Pr(\tau_i = 2) = \frac{1}{2} \lambda_i^2 e^{-\lambda_i} \quad (2)$$

This study further assumes that major incidents occurring on different freeway segments are not correlated. Because of the potential of having secondary incidents, this assumption may not hold for all road segments. However, previous research has revealed that the identification of secondary incidents is extremely difficulty because of the lack of data (Chou and Miller-Hooks 2010; Zhang and Khattak 2010). Because the secondary incident rate is insignificant (approximately 2–3%) based on the empirical data [Chang and Rochon (2009), which used the data recorded in accident reports from the Maryland State Police (MSP) for their estimation], the independence assumption will be maintained in this research, and future studies will further investigate correlations in incident probability. The probability of two incidents happening on segment i and k during the same time period is given by

$$\Pr(\tau_i = 1, \tau_k = 1) = \begin{cases} \Pr(\tau_i = 1) \times \Pr(\tau_k = 1), & \text{if } i \neq k \\ \Pr(\tau_i = 2), & \text{otherwise} \end{cases} \quad (3)$$

Given these inputs, a model for allocating emergency response units should provide

- A response unit location strategy Y , as follows:

$$y_j = \begin{cases} 1, & \text{if } j \in M \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

- An allocation strategy, denoted by X , where each member x_{ij} equals 1 if accidents at segment i will be taken care by response units located at location j , and 0 otherwise.

Differing from the previous model (Zhu et al. 2012), this study also considers the possibility that two incidents (i, k) happen at the same time and a back-up strategy Z must be designed. Each member z_{ikj} equals 1 if the second incident k is taken care of by response units located at location j , given the first incident occurring at the location i , and 0 otherwise. Because of the extremely low probability, this study ignores the cases in which more than two incidents occur concurrently.

The objective of traffic management agencies can then be summarized as follows:

$$\min W = \sum_i \sum_j \Pr(\tau_i = 1) x_{ij} t_{ij} + \sum_i \sum_k \sum_j \Pr(\tau_i = 1, \tau_k = 1) \times (x_{ij} t_{ij} + z_{ikj} t_{kj}) \quad (5)$$

The first term of the objective function represents the expected response time according to the allocation strategy X if only one accident happens. The second term represents the expected additional response time if a second incident happens at the same time. The strategy of dispatching the second response unit to the second incident is summarized in Z , which will be determined by the optimization algorithm. The overall objective is to minimize the total expected response time W .

The following constraints are applied for both the demand and the supply sides: (1) every freeway segment i and k must be served; (2) response units can only be dispatched from location j if they are stationed there ($y_j = 1$); (3) if the only available response unit at location j has been dispatched to take care the incident at location i , it cannot be dispatched to respond to the second incident at location k ; and (4) the total number of available response units is limited by the available resources (P). Constraint 1 can be expressed as

$$\sum_j x_{ij} = 1 \quad \text{for all } i \quad (6)$$

$$\sum_j z_{ikj} = 1 \quad \text{for all } i, k \quad (7)$$

Constraint 2 is formulated as

$$y_j = \begin{cases} 1, & \exists x_{ij} = 1 \text{ or } \exists z_{ikj} = 1 \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

By introducing a large number L (which should be equal to or larger than I), constraint 2 can be reformulated as

$$\sum_i x_{ij} \leq L y_j \quad \text{for all } j \quad (9)$$

$$\sum_i \sum_k z_{ikj} \leq L y_j \quad \text{for all } j \quad (10)$$

Constraint 3 is formulated as

$$x_{ij} + z_{jkj} \leq 1 \quad \text{for any } i, k, j \quad (11)$$

Constraint 4 can simply be expressed as

$$\sum_j y_j \leq P \quad (12)$$

Given the aforementioned assumptions, the allocation task becomes a p -median problem with both x_{ij} and y_j as binary variables,

as follows: $x_{ij} \in 0 \cup 1, \forall i, j; z_{ikj} \in 0 \cup 1, \forall i, k, j;$ and $y_j \in 0 \cup 1, \forall i, j.$

The models can then be applied to decide the optimal locations of freeway incident response vehicles, given the empirical distribution of accidents on each freeway segment and the corresponding travel cost matrix. To facilitate the understanding of the model, the major assumptions are summarized as follows:

- Incidents occur following a Poisson distribution with the mean and variance empirically estimated for each freeway segment and for each time period based on historical data;
- The covariance of incidents on different freeway segments was not considered;
- No more than two incidents can happen during the same period;
- Each incident has to be responded to by one and only one response unit;
- In the case in which two incidents happen during the same time period, two different response units have to be dispatched; and
- The travel speed of response units is assumed constant.

The proposed model is solved by a *Cplex* algorithm in an *ILOG* environment. The dispatch model will generate the response time from dispatch until arrival under the optimized response unit location-allocation strategy. In Maryland, CHART defines the start of an incident response by either the detection of an incident by a patrol unit or by the reception of an incident report at any operation center. The total response time using prepositioning and dispatching strategies will be the sum of travel time of the response team, which are outputs from this model, and the incident detection and response units dispatching time. The latter depends on the availability of traffic monitoring devices, the efficiency of detection and communication, the staffing level, and the efficiency of responsible agencies. The availability of a response unit is also affected by the clearance time, which is the period between the arrival and the departure of the response unit. Depending on the severity, the response unit may stay at the site until all traffic lanes are completely cleared, or move on if it is a minor incident. The response unit has the discretion to decide how to respond if the operation center requests an urgent assist in other sites. Therefore, the impact of clearance time on availability of response units varies from case to case. Limited by its scope, this study does not consider the impact of clearance time on availability of response units. Incident detection and dispatching time will be empirically estimated for different time periods using local data. The results will be presented and discussed in the "Illustration of the Study Site" section.

Dispatching Model with Reliability Constraints

As discussed in the introductory section, traffic agencies usually deploy incident management systems based on history incident data. Considering the significant variance in the distribution of incident frequency, traffic agencies may want to impose a constraint on the total response variance so as to insure a certain performance level in the worst-case scenario. An empirical study on variance in incident rates will be presented in the next section. Given the average incident frequency at section i during a period $\lambda_i \in N(\mu_i, \sigma_i^2)$, a new constraint shall be imposed, as follows:

$$\text{var}(W) \leq v_c \quad (13)$$

where v_c = critical value preselected by traffic management agencies.

Because the variance of total response time includes the second-order term of decision variables, this problem may look like a non-linear programming problem. However, notice that

$$x_{ij} \times x_{ij} = x_{ij} \quad \text{for any } i, j \quad (14)$$

Similarly, there is

$$z_{ikj} \times z_{ikj} = z_{ikj} \quad \text{for any } i, k, j \quad (15)$$

Further

$$x_{ij} \times z_{ikj} = 0 \quad \text{for any } i, k, j \quad (16)$$

This can be easily confirmed from Eq. (11).

Therefore, Eq. (13) can be converted into a linear constraint if it is assumed that incidents occurring on different segments are independent

$$\begin{aligned} \text{var}(W) &= \sum_i \sum_j \text{var}[\text{Pr}(\tau_i = 1)] x_{ij} t_{ij}^2 \\ &+ \sum_i \sum_k \sum_j \text{var}[\text{Pr}(\tau_i = 1, \tau_k = 1)] \\ &\times (x_{ij} t_{ij}^2 + z_{ikj} t_{kj}^2) \leq v_c \end{aligned} \quad (17)$$

Because the average incident rate for all subsegments of the highway network in theory follows a normal distribution based on the central limit theorem (CLT), the objective function should be converted as minimizing the expected total response time

$$\begin{aligned} \min W &= \sum_i \sum_j E[\text{Pr}(\tau_i = 1)] x_{ij} t_{ij} \\ &+ \sum_i \sum_k \sum_j E[\text{Pr}(\tau_i = 1, \tau_k = 1)] \\ &\times (x_{ij} t_{ij} + z_{ikj} t_{kj}) \end{aligned} \quad (18)$$

To solve this problem, one has to derive the expectation and variance of incident probabilities, given the distribution of the average incident frequency on each segment derived from the historical data. For convenience of presentation, the subscript for the segment from the presentation is hereafter ignored

$$\lambda \sim N(\mu, \sigma^2) \quad (19)$$

$$f(\lambda) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(\lambda - \mu)^2}{2\sigma^2}\right) \quad (20)$$

$$\text{Pr}(x = 1) = \lambda e^{-\lambda} \quad (21)$$

$$\begin{aligned} E[\text{Pr}(x = 1)] &= \int_{-\infty}^{\infty} \lambda e^{-\lambda} f(\lambda) d\lambda \\ &= \int_{-\infty}^{\infty} \lambda e^{-\lambda} \frac{1}{\sqrt{2x\sigma}} \exp\left[-\frac{(\lambda - \mu)^2}{2\sigma^2}\right] d\lambda \\ &= \exp\left(-\mu + \frac{\sigma^2}{2}\right) (\mu + \sigma^2 - \sigma^2) \\ &= \mu \exp\left(-\mu + \frac{\sigma^2}{2}\right) \end{aligned} \quad (22)$$

$$\begin{aligned} \text{Var}[\text{Pr}(x = 1)] &= \int_{-\infty}^{\infty} \{\lambda e^{-\lambda} - E[\text{Pr}(x = 1)]\}^2 f(\lambda) d\lambda \\ &= \exp(-2\mu + \sigma^2) \{\sigma^2 \exp(\sigma^2) + \mu^2 [\exp(\sigma^2) - 1]\} \end{aligned} \quad (23)$$

Eqs. 22 and 23 derived the closed-form solutions for the expectation and variance of the probability that an incident is going to happen on a freeway segment during a certain time period. The results for the scenarios in which two incidents occur concurrently become quite complex. To simplify the process, this research uses the Monte Carlo simulation approach to derive the expectation and variance of corresponding terms. For the crossing

terms, one should notice that if A and B are independent, then

$$\text{Var}(AB) = E^2(A)\text{Var}(B) + E^2(B)\text{Var}(A) + \text{Var}(A)\text{Var}(B) \quad (24)$$

Therefore, one can derive all unknowns in Eq. (17) and the objective function Eq. (18) based on the historical data, then solve the problem with a reliability constraint Eq. (17) and compare it with the base-case scenario. Numerical evaluation results will be presented in the section of “Strategic Alternatives and Study Results”.

Patrolling Model

Note that the total incident response time usually includes three components: (1) incident detection time; (2) waiting time (from incident detection until a response unit is dispatched); and (3) response time from dispatch until arrival (Hall 2002). The detection of incidents during the daytime is relatively convenient, and the dispatching process is usually faster because of better coordination and more staff during the working hours. Therefore, a dispatching system may exhibit some advantage during peak periods. In contrast, the first two terms in the total response time may become significantly longer late at night, making the patrolling system more preferable because it can potentially minimize the response time.

This study considers a patrolling system that divides the entire freeway network into N beats. It is assumed that set A includes all link segments that belong to beat n , and there are v_n vehicles patrolling on this segment. Thus, the average headway becomes $\sum_{i,j \in A} t_{ij}/v_n$. The average detection time for any incident is half of the headway $\frac{1}{2} \sum_{i,j \in A} t_{ij}/v_n$. The objective to minimize the total detection time can be formulated as follows:

$$\min W = \sum_n \frac{1}{2} \sum_{i,j \in A_n} t_{ij} \Pr(\tau_{ij} = 1)/v_n \quad (25)$$

A generalized design for a freeway patrol service system is recognized to be very complex, and in many cases is formulated as a mixed-integer nonlinear optimization problem that does not have an efficient solution algorithm. The focus of this paper is not to design such a system, but to compare the dispatching with the patrolling systems. Therefore, this study will not further discuss its solution algorithm. Instead, the numerical example provided in this study will take advantage of the network configuration of the beltway and solve the problem heuristically by the following steps:

1. Calculate the sum of $t_i \Pr(\tau_i = 1)$ on the entire beltway.
2. Divide the sum by n , the number of available response units.
3. Starting from the rightmost point of I-495 where it hits the Virginia boundary, patrol beats are built one-by-one with the equal incident headway $t_i \times \Pr(\tau_i = 1)/n$. Because a patrol vehicle cannot patrol only part of a freeway segment between two exits (this study does not consider the probability of emergency crossings on the median), the segment allocations are rounded.
4. Calculate the total response time on each beat A by applying the average detection time, $\sum_{i,j \in A} t_{ij}$, to all incidents that occurred on beat A .

On a beltway, solutions from this heuristic are consistent with the minimization problem described by Eq. (25). The next section will present the data that are used in this study. The “Strategic Alternatives and Study Results” section will then apply the models proposed in this study along with analysis results.

Illustration of the Study Site

Assuming that the demand (incidents) and the supply (response units) are located on nodes (freeway exits), one can then compute the travel time from node i to node j based on the link travel time described in the next subsection. This assumption is for convenience of presentation and can be easily relaxed. Incident data for the case study are obtained from the CHART II database and the Maryland Accident Analysis Reporting System (MAARS), whereas travel speed on each freeway segment was collected from INRIX detectors and documented by the Center for Advanced Transportation Technology (CATT) laboratory at the University of Maryland.

To conduct the performance evaluation, different models are first computed by using data from 2006–2008, and then evaluated with 2009 and 2010 data. The incident probability on freeway segment i , $\Pr(\tau_i = 1)$, is estimated based on the data from 2006–2008.

Study Site and Current Operations Status

The study site is the Capital Beltway (I-495/I-95) in Maryland from the Woodrow Wilson Bridge to the American Legion Bridge (see Fig. 2). It is a 68-km (42-mi) long segment with 30 distinct exits and is one of the major corridors managed by CHART.

Currently, Traffic Operation Center-3 (TOC-3) of CHART is located at exit 25, having 11 field operations units to manage incidents occurring on I-95, I-270, US-50, MD-295, and the Capital Beltway in Prince George’s and Montgomery counties. They provide incident response and driver assists for 16 h each day (5 a.m.–9 p.m.) Monday–Friday.

Fig. 3 summarizes the average speed by time of day in 2009. It indicates that the average traffic speed varies over the locations, the directions, and the time of day. For the inner loop of I-495/I-95, the average speeds for the p.m. peak period fluctuate considerably over these exits, whereas they display relatively constant patterns during other times of day over the entire network. In contrast, the outer loop of I-495/I-95 exhibits quite stable speed patterns from exits 0–25, and the speed drops significantly beyond that segment. Notably, during the a.m. peak period, its speed is below 48 km/h (30 mi/h) between exits 27 and 28, although the average speed at night at the same segment is approximately 96 km/h (60 mi/h). Therefore, travel times for response units may vary significantly, depending on the travel direction and the time of day. Thus, the travel time used in this study is asymmetric for each time interval.

Incident Rates

Fig. 4 illustrates different patterns for average incident rates by time of day along the Capital Beltway in 2011. The CHART defines incidents as any events that affect traffic flow on the roadway. These include disabled vehicles on the roadway/shoulder, vehicle fire, road debris, emergency roadwork, police activities, and vehicle crashes. However, if a minor collision was resolved by drivers and was not reported to CHART, it would not be recorded in the CHART database and cannot be considered in this study. Because of the limit in scope, this study does not differentiate incidents with different levels of severity. However, the severity could have a significant impact on the clearance time, which would be addressed in future studies.

The figure shows significant fluctuations of the average incident rate along the study links. Also, each time of day shows a unique pattern of its incident rate distribution. It is obvious that most segments during p.m. peak hours have the highest average incident

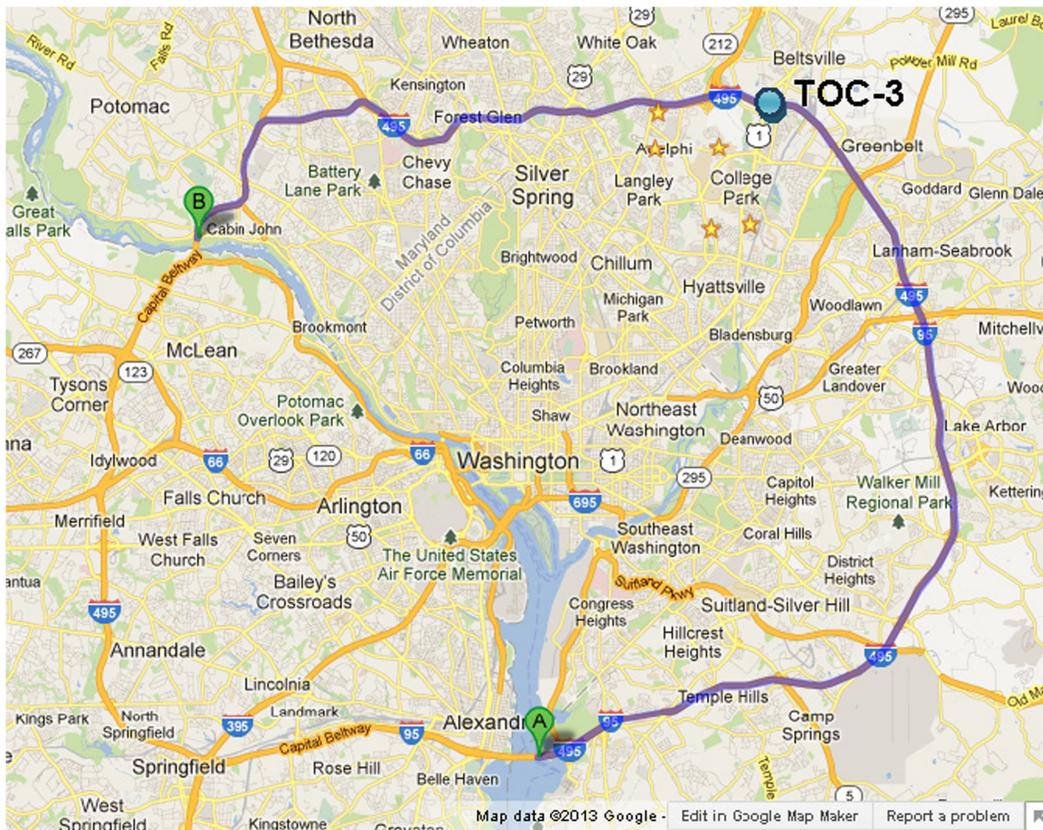


Fig. 2. Study links in the Capital Beltway (I-495/I-95) in Maryland (Map data © 2013 Google)

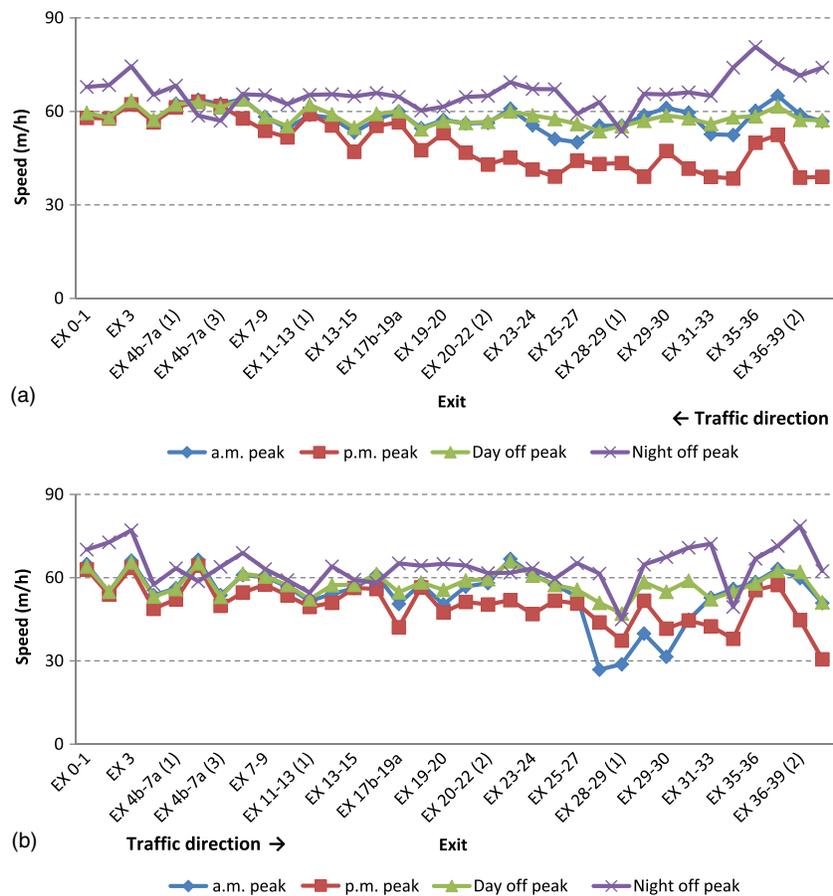


Fig. 3. Average traffic speed patterns by time of day on I-495/I-95: (a) inner loop; (b) outer loop

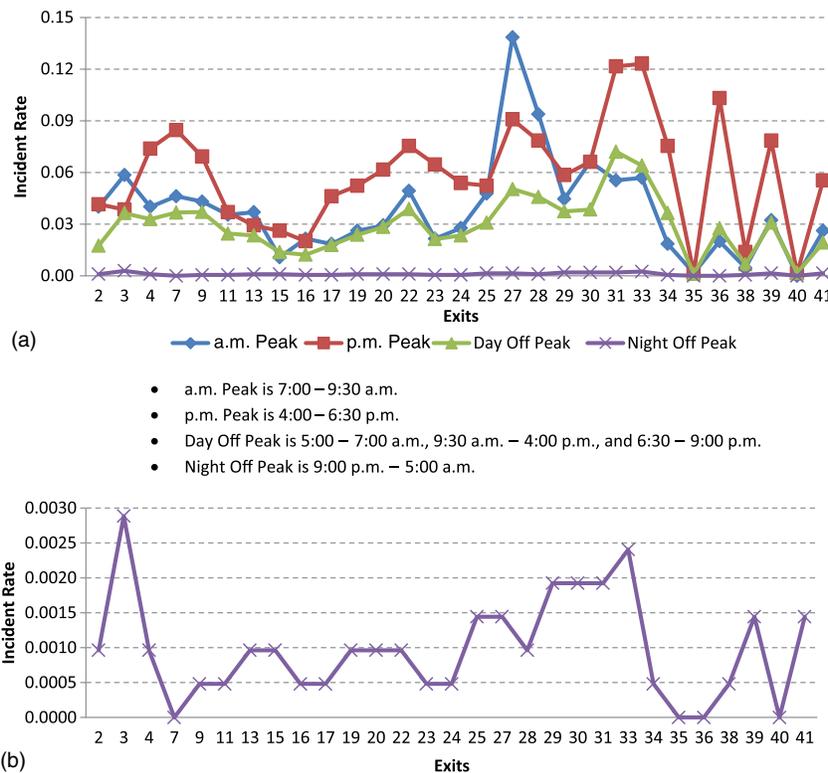


Fig. 4. Average incident rates by time of day on the Capital Beltway (I-495/I-95), Maryland: (a) different time of day; (b) night only

rate, whereas the incident rates at night are much lower than those at any other times, as shown in the vertical scale of Fig. 4(b). Historical field data from MAARS indicate that approximately 10% of total incidents (278 cases in the Capital Beltway in 2011) occurred at night, and most of these cases were collision-type incidents. The locations showing the most frequent incidents are also different over different times of a day. For example, exits 27, 33, 31, and 3 show the highest incident rate during the a.m. peak, p.m. peak, off-peak at daytime, and night, respectively.

These findings would be valuable for responsible agencies intending to take the experience-based strategy to allocate emergency response units at these sites having high incident rates. Moreover,

Fig. 4 shows that traffic operators should consider the distributions of incident rates both by the segment and the time of day to determine the optimal fleet size for each shift and to allocate available resources so as to improve the efficiency and effectiveness of an incident response program.

Response Efficiency

According to the Highway Capacity Manual [Transportation Research Board (TRB) 1994], incident durations are defined to be the time elapsed from the onset of an incident to the end of its clearance. In most cases, it is impossible to know exactly when

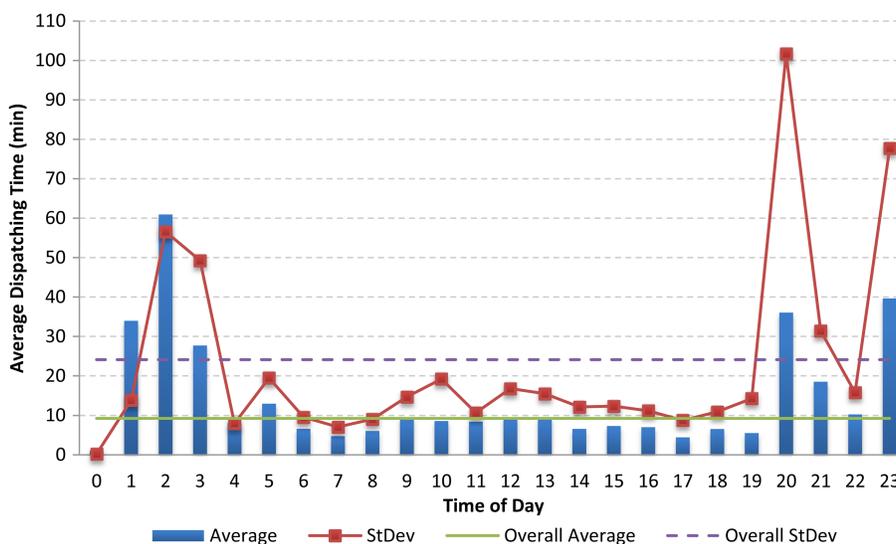


Fig. 5. Distribution of average dispatching time at different time of the day based on 2009 data

Table 2. Assigned Locations/Segments for Available Response Units by Strategy

Number of units available	a.m. peak		Night	
	Dispatching (prepositioned locations in exit #)	Patrolling (starting and ending exit # of each beat)	Dispatching (prepositioned locations in exit #)	Patrolling (starting and ending exit # of each beat)
2	15, 30	(1, 24), (24, 41)	4, 25	(1, 15), (15, 41)
3	4, 23, 31	(1, 15), (15, 27), (27, 41)	3, 19, 31	(1, 7), (7, 24), (24, 41)
4	4, 20, 29, 34	(1, 7), (7, 23), (23, 28), (28, 41)	3, 15, 24, 33	(1, 4), (4, 17), (17, 29), (29, 41)
5	3, 11, 23, 30, 35	(1, 4), (4, 19), (19, 27), (27, 31), (31, 41)	3, 9, 17, 25, 34	(1, 3), (3, 7), (7, 17), (17, 27), (27, 41)
6	3, 9, 19, 24, 30, 35	(1, 4), (4, 17), (17, 25), (25, 27), (27, 30), (30, 41)	3, 9, 16, 24, 31, 38	(1, 3), (3, 7), (7, 15), (15, 23), (23, 30), (30, 41)
7	3, 9, 19, 24, 30, 33, 40	(1, 4), (4, 15), (15, 23), (23, 27), (27, 30), (30, 36), (36, 41)	2, 4, 9, 16, 24, 31, 38	(1, 3), (3, 4), (4, 11), (11, 20), (20, 27), (27, 31), (31, 41)
8	3, 9, 16, 20, 25, 30, 33, 40	(1, 3), (3, 4), (4, 13), (13, 20), (20, 25), (25, 27), (27, 29), (29, 41)	2, 4, 9, 16, 19, 24, 31, 38	(1, 2), (2, 3), (3, 4), (4, 9), (9, 17), (17, 23), (23, 29), (29, 41)
9	3, 9, 16, 20, 23, 27, 30, 33, 40	(1, 3), (3, 4), (4, 11), (11, 20), (20, 25), (25, 27), (27, 29), (29, 31), (31, 41)	2, 4, 9, 16, 19, 23, 27, 31, 38	(1, 2), (2, 3), (3, 4), (4, 9), (9, 16), (16, 22), (22, 28), (28, 31), (31, 41)
10	2, 4, 9, 16, 20, 23, 27, 30, 33, 40	(1, 3), (3, 4), (4, 11), (11, 19), (19, 23), (23, 25), (25, 27), (27, 28), (28, 30), (30, 41)	1, 3, 4, 9, 16, 19, 23, 27, 31, 38	(1, 2), (2, 3), (3, 4), (4, 9), (9, 15), (15, 20), (20, 25), (25, 30), (30, 33), (33, 41)
11	2, 4, 9, 16, 20, 23, 27, 28, 30, 34, 40	(1, 3), (3, 4), (4, 9), (9, 17), (17, 22), (22, 25), (25, 27), (27, 28), (28, 30), (30, 33), (33, 41)	1, 3, 4, 9, 16, 19, 23, 27, 31, 34, 40	(1, 2), (2, 3), (3, 4), (4, 9), (9, 15), (15, 19), (19, 24), (24, 28), (28, 31), (31, 38), (38, 41)

Note: a.m. peak is from 7:00–9:30 a.m., and night is from 9 p.m. to 5 a.m.

the detected incident occurred; therefore, many studies defined the incident duration as the time interval between the incident detected and cleared (e.g., Garib et al. 1997; Nam and Mannering 2000; Smith and Smith 2001). This study focuses on the total incident response time, which is the sum of the dispatching time and the travel time of the response team. The dispatching time represents a significant portion of the total response time in the detection and dispatching system, and its length varies during different periods.

Fig. 5 shows that the average dispatching times between 8 p.m. to 3 a.m. are much longer than those during other time periods. The variances of dispatching times at night also seem much higher than those during daytime. The main reason could be the difference in the service quality of the response team. Most incident management units operate during the daytime so that incidents are more likely to be managed promptly and efficiently. In contrast, the primary CHART operation centers (TOCs 3, 4, and 7) currently do not operate during the night so that the overall incident response efficiency was significantly degraded over those periods.

Strategic Alternatives and Study Results

Results for Dispatching and Patrolling Strategies

To evaluate the performance of the proposed model, this study uses the annual total response time as the measure of effectiveness for comparison. Scenarios with different available units (2–11) are tested here, given the maximum of 11 units for patrol operations under CHART. Table 2 summarizes the optimized operation strategies. The numbers under patrolling strategies represent the locations (freeway segment or exit number) where the response units should be prepositioned, whereas the numbers under roaming strategies represents the starting point and ending point (represented using the corresponding freeway segment or exit numbers) of each roaming beat. The number of prepositioned locations or the number of beats should equal the number of available units under the optimized condition.

These candidate strategies are applied to the 2009 incident data for comparison of response efficiency. The numbers of incidents for this analysis are 1,167 and 78 for the a.m. peak and night periods, respectively, in 2009.

Fig. 6 displays the estimated total response time by allocating the emergency team according to each strategy. During the a.m. peak period, the dispatching strategy outperforms the patrolling strategy for most fleet sizes, especially when only small numbers of units are available. However, the total response time by patrolling strategies drops drastically as the fleet size increases, and eventually outperforms the dispatching strategy when given 10 units. In contrast, the patrolling strategy consistently outperforms the dispatching strategy for any fleet sizes during the night, and their differences in the total travel time are consistently significant.

These results imply that during the peak periods, dispatching strategies would perform better than patrolling strategies when only a small number of units is available. It could be because incident detection efficiency by various sources (e.g., CCTV, media, citizens, and police) is fairly good during peak hours. Thus, response times by dispatching strategies are likely to be shorter than those by the patrolling operations. However, as more units are available, their efficiencies evolve to the same level, and at a certain point, the patrolling strategy would work better than the dispatching method. This finding can be explained by the fact that as more units are available, their assigned segments become shorter and the frequency of patrolling increases so that their chances to find incidents

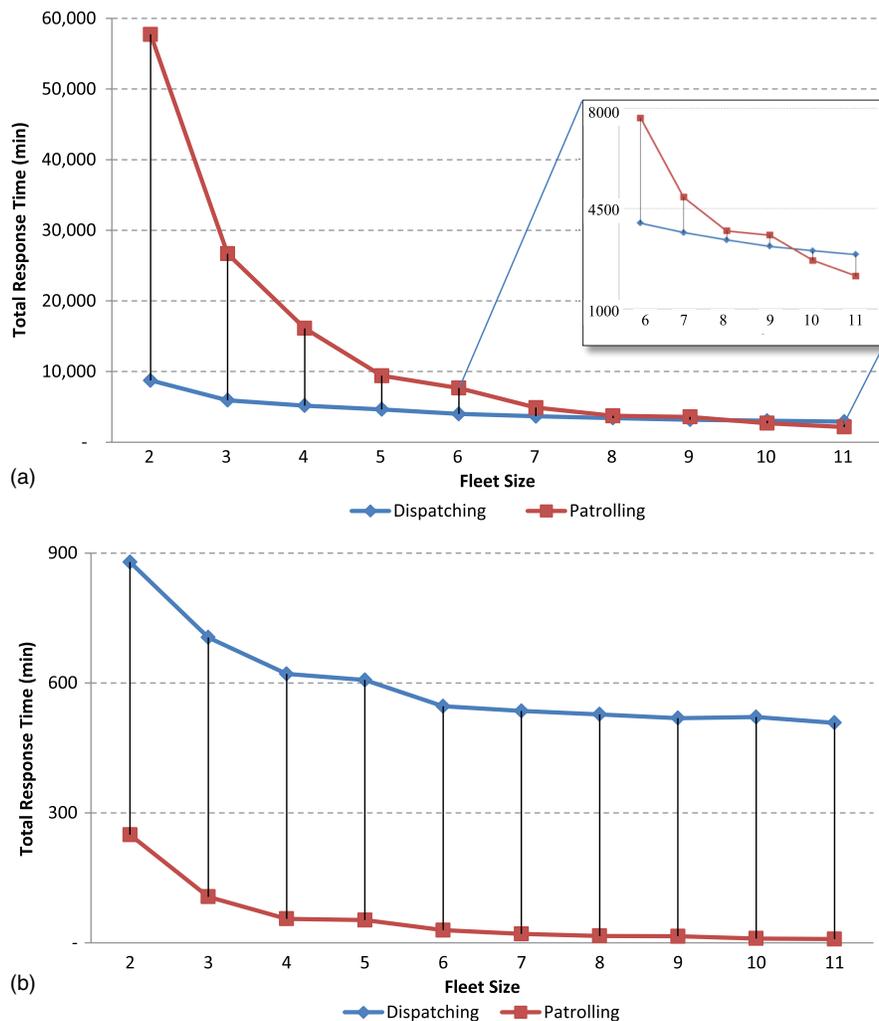


Fig. 6. Total response times by strategy: (a) a.m. peak; (b) night

would increase and the travel times to reach incident sites could be shorter than those by dispatching from depots. Obviously, the detection times at night would be much longer because of the limited availability of detection sources. Therefore, patrolling strategies would save a significant amount of detection times.

Results with Variances of Incident Probabilities

Taking into account variances of incident probabilities has slightly changed the optimal allocations for response units under the dispatching strategy. For example, the proposed model, considering the variance of incident probabilities, yields the following five locations for deploying five response units during the a.m. peak hours: exits 3, 13, 23, 30, and 35. Similarly, the model, considering the variance, suggests exits 3, 13, 23, 30, and 38 for optimal allocation during a night period. As presented in Table 3, the annual

Table 3. Comparisons of Total Travel Time and Optimal Locations for w/ and w/o Variance Consideration

Time of day	w/o variance		w/ variance	
	Total travel time (min)	Optimal location	Total travel time (min)	Optimal location
a.m. peak	2,545	3, 11, 23, 30, 35	2,534	3, 13, 23, 30, 35
Night	168	3, 9, 17, 25, 34	134	3, 13, 23, 30, 38

total travel time (using 2009 data) for allocating response units, based on the models accounting for data variances, are slightly better than those without using such information. This indicates that the variances of incident patterns would also be an important factor in designing an incident response strategy.

Optimal Response Units from the Benefit-Cost Perspective

The aforementioned analysis provides the optimal deployment strategy for the available emergency response units. However, in view of the socioeconomic cost and the diminishing resources for traffic management, this study further explores the optimal number of emergency response units under the projected distribution of incidents and traffic patterns. The socioeconomic cost reduction because of an efficient incident response strategy includes the savings on incident-induced delays, fuel consumption, and emissions.

Various formulas have been proposed in the literature to estimate delays and the associated benefits (Maccubbin et al. 2008; Chang and Rochon 2009; Lindley 1989; Latoski et al. 1998; Chou et al. 2010; Bertini 2006; Haghani et al. 2006; Guin et al. 2007). To be consistent with the practice of the Maryland State Highway Administration (MSHA) and take advantage of the locally calibrated parameters, this research adopts the following

Table 4. Traffic and Incident Information Used to Estimate Incident-Induced Delays

Variables	a.m. peak	Night
AADT on I-495	195,238	195,238
PHF	0.95	N/A
Adjusted AADT for peak hours	205,513	205,513
Number of lanes (one direction)	4	4
Average clearance time with CHART (min)	12.05	33.00
Average dispatching time with CHART (min)	1.79	5.62
Average lane blockage attributable to incidents	1	1

procedure for total delay and direct benefit estimation (Chang and Rochon 2009).

1. Compute the total delay reduction with each candidate strategy using the following formulation and information:

$$D = \sum_i e^{\mu} \times f^{\varphi} \times \left(\frac{b}{n}\right)^{\theta} \times d^{\gamma} \times N \quad (26)$$

where D = total excessive delay incurred by the incidents on top of the recurrent congestion; f = traffic volume (vehicle per lane per h) at the segment i ; b and n = number of lanes blocked and the total number of lanes, respectively; d = average incident duration (h) at the segment i ; N = total number of incidents at the segment i ; and $\mu, \varphi, \theta, \gamma$ = parameters. In this study, $-10.19, 2.8, 1.4,$ and 1.78 have been used, respectively, as recommended by CHART.

Table 4 summarizes the value of parameters used in this study to estimate the savings on delay and other benefits related to different incident response strategies.

2. Estimate the fuel consumption and emissions from the delay reduction using the following information:
 - Fuel consumption for passenger cars: 0.59 L (0.156 gal.) of gasoline/h of delay (Ohio Air Quality Development Authority)
 - HC: 13.073 g/h of delay (provided by MDOT in 2000)
 - CO: 146.831 g/h of delay (provided by MDOT in 2000)
 - NO: 6.261 g/h of delay (provided by MDOT in 2000)
 - CO₂: 8.8 kg (19.56 lbs) CO₂/gal. of gasoline (Energy Information Administration)
3. Convert the saved delay, fuel, and emissions to the monetary value.

Similar to step 3, monetary conversion factors have been used to estimate the reduced delay and associated by-products in a monetary value. The values and sources for factors are shown as follows:

- Delay: \$27.37/h (U.S. Census Bureau in 2008)
- Fuel: \$2.32/gal. (Energy Information Administration in 2009)
- HC: \$6,700/t (DeCorla-Souza et al. 1998)
- CO: \$6,360/t (DeCorla-Souza et al. 1998)
- NO: \$12,875/t (DeCorla-Souza et al. 1998)
- CO₂: \$23/metric t [Congressional Budget Office (CBO)'s cost estimate for S. 2191, America's Climate Security Act of 2007]

To analyze the marginal contribution and cost of each additional response unit, this study evaluated the emergency response team of 2–11 units. Because dispatching strategies outperform patrolling strategies at most fleet sizes during the a.m. peak hours, the

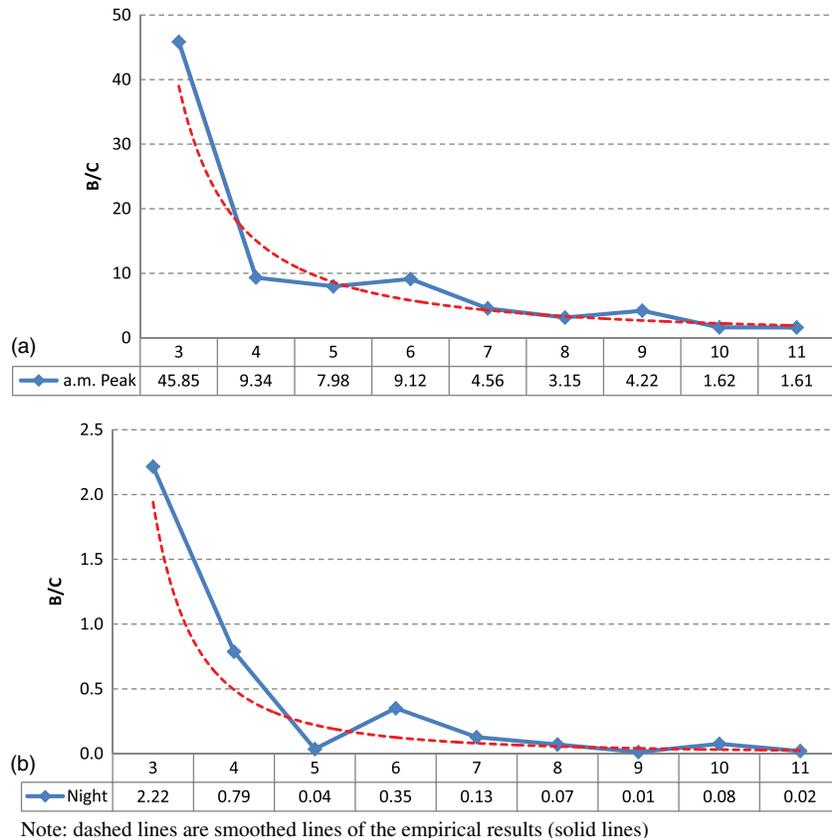


Fig. 7. Marginal benefit-cost ratios for additional response unit

marginal benefit-cost analysis was conducted only for the dispatching strategy. By the same token, the marginal benefit/cost analysis was conducted only for the patrolling strategy for the night period. As shown in Fig. 7, the marginal benefit-cost ratios for both analyses quickly decrease as the number of response unit increases, and then their slopes stay relatively flat. The information provided by this analysis with the smooth dashed line can serve as the basis for traffic management agencies to decide the most cost-beneficial fleet size.

Conclusions

Both the patrolling and the prepositioned strategies for allocating emergency traffic response units have been implemented in practice, and each has its strengths and limitations. There has been no consensus in the literature about the superiority of one over the other, and the best practice may well depend on the incident patterns and traffic conditions. With a limited budget and resources, an analysis tool is needed to support the analysis, design, and implementation of the most cost effective incident response system. This study has proposed a robust model to optimize the application of the dispatching strategy with available response units. An efficient heuristic has also been proposed to best use the patrolling strategy that takes advantage of the target traffic corridor's unique geometric features. Performance of the two different systems has been evaluated and compared with various incidents and various patterns.

Because the assumption of incident patterns has a critical impact on model performance, this study has used the empirical data and statistical tests to select the best fit distribution. For most road segments, a Poisson distribution fits the data well. It also exhibits a strong advantage in simplifying the modeling process and keeps the results tractable. Therefore, this study adopts a Poisson distribution to model incident patterns.

The strength of the proposed model has been demonstrated through an application on the I-495/I-95 Capital Beltway in the metropolitan Washington, DC area. Incident data from 2006–2010 were used. The extensive experimental results have revealed that the effectiveness of the two different response strategies varies with some critical factors, including the spatial distribution of incident frequency over different times of a day, the fleet size of the response team, the congestion level, and the available detection sources. The analysis results clearly reveal that the dispatching strategy is more preferable during the peak periods, whereas the patrolling strategy leads to a shorter response time during the night under the traffic and incident conditions observed on the Capital Beltway. The reason is that the dispatch time during the night is much longer for the TOCs in Maryland. This pattern could be true for other metropolitan areas given the challenge of detecting an incident and the low level of available staff at night. However, the model has to be calibrated using locally collected data applied before such conclusions can be drawn.

The paper has further reported that the variance of incident rates may influence the performance of incident response systems. Its impact may become significant if the incident patterns are volatile and a limited number of response units are available. Because traffic agencies usually deploy their system based on historical data, the impact of variance in incident rates has to be carefully evaluated and fully considered in system design.

Analysis in this study demonstrates that the proposed methods can support the responsible highway agencies to best operate their available incident response fleet for different time periods based on all resources, the quality of available historical incident data, and

various incident detection sources. The study has further presented a methodology to determine the most cost-beneficial fleet size operated with the proposed strategies, considering the marginal cost and benefit of an additional response unit on the resulting total social benefits. The analysis results with the Capital Beltway, despite being exploratory in nature, could serve as the basis for highway agencies to review and optimize their incident response and management program.

References

- Bertini, R. (2006). "You are the traffic jam: An examination of congestion measures." *85th Annual Meeting of the Transportation Research Board*, Transportation Research Board, Washington, DC.
- Chang, G.-L., and Rochon, S. (2009). "Performance evaluation and benefit analysis for CHART." *Technical Rep.*, Maryland State Highway Administration, Hanover, MD.
- Chou, C., Miller-Hooks, E., and Promisel, I. (2010). "Benefit-cost analysis of freeway service patrol programs: Methodology and case study." *Adv. Transp. Stud. Int. J. Sec. B*, 20, 81–96.
- Chou, C., and Miller-Hooks, E. (2010). "Simulation-based secondary incident filtering method." *J. Transp. Eng.*, 10.1061/(ASCE)TE.1943-5436.0000134, 746–754.
- Church, R., and ReVelle, C. (1974). "The maximal covering location problem." *Pap. Reg. Sci.*, 32(1), 101–118.
- DeCorla-Souza, P., Cohen, H., Haling, D., and Hunt, J. (1998). "Using STEAM for benefit-cost analysis of transportation alternatives." *Transportation Research Record 1649*, Transportation Research Board, Washington, DC, 63–71.
- Federal Highway Administration (FHWA). (2005). "An initial assessment of freight bottlenecks on highways." *FHWA white paper*, Washington, DC.
- Friendly, M. (2000). *Visualizing categorical data*, SAS Institute, Cary, NC.
- Garib, A., Radwan, A., and Al-Deek, H. (1997). "Estimating magnitude and duration of incident delays." *J. Transp. Eng.*, 10.1061/(ASCE)0733-947X(1997)123:6(459), 459–466.
- Geroliminis, N., Karlaftis, M. G., and Skabardonis, A. (2009). "A spatial queuing model for the emergency vehicle districting and location problem." *Transp. Res. Part B: Method.*, 43(7), 798–811.
- Guin, A., Porter, C., Smith, B., and Holmes, C. (2007). "Benefits analysis for incident management program integrated with intelligent transportation systems operations-case study." *J. Transp. Res. Board*, 2000(1), 78–87.
- Haghani, A., Iliescu, D., Hamed, M., and Yang, S. (2006). "Methodology for quantifying the cost effectiveness of freeway service patrols programs—Case study: H.E.L.P. program." *Technical Rep.*, Univ. of Maryland, College Park, MD.
- Hakimi, S. (1964). "Optimal locations of switching centers and the absolute centers and medians of a graph." *Oper. Res.*, 12(3), 450–459.
- Hall, R. (2002). "Incident dispatching, clearance, and delay." *Transp. Res. Part A*, 36(1), 1–16.
- Kweon, Y., and Kockelman, K. (2003). "Overall injury risk to different drivers: Combining exposure, frequency, and severity models." *Accid. Anal. Prev.*, 35(4), 441–450.
- Larson, R., and Odoni, A. (1981). *Urban operations research*, Prentice-Hall, Englewood Cliffs, NJ.
- Latoski, S., Pal, R., and Sinha, K. (1998). "An evaluation of the cost effectiveness of the Hoosier helper program and framework for the design of ITS optimal system configuration, phase 1." *Paper 346, FHWA/IN/JTRP-97/09*, Joint Transportation Research Program, Indiana Dept. of Transportation and Purdue Univ., West Lafayette, Indiana.
- Lindley, J. (1989). "The urban freeway congestion problem." *Transportation Research Circular 344*, Transportation Research Board (TRB), National Research Council, Washington, DC.
- Lord, D., and Mannering, F. (2010). "The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives." *Transp. Res. Part A: Pol. Pract.*, 44(5), 291–305.

- Lord, D., Washington, S., and Ivan, J. (2005). "Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: Balancing statistical fit and theory." *Accid. Anal. Prev.*, 37(1), 35–46.
- Lou, Y., Yin, Y., and Lawphongpanich, S. (2010). "Freeway service patrol deployment planning for incident management and congestion mitigation." *Transp. Res. Part C Emerg. Technol.*, 19(2), 283–295.
- Ma, J., Kockelman, K., and Damien, P. (2008). "A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods." *Accid. Anal. Prev.*, 40(3), 964–975.
- Maccubbin, R., et al. (2008). "Intelligent transportation systems benefits, costs, deployment, and lessons learned: 2008 update." *FHWA-IPO Rep. 08-032*, Noblis, Washington, DC.
- Nam, D., and Mannering, F. (2000). "An exploratory hazard-based analysis of highway incident duration." *Transp. Res. Part A*, 34A(2), 85–102.
- Ozbay, K., and Bartin, B. (2003). "Incident management simulation." *Simulation*, 79(2), 69–82.
- Pal, R., and Bose, I. (2009). "An optimization based approach for deployment of roadway incident response vehicles with reliability constraints." *Eur. J. Oper. Res.*, 198(2), 452–463.
- Pal, R., and Sinha, K. (1997). "A framework for locating highway incident response vehicles in urban areas." *INFORMS National Meeting*, Institute for Operations Research and the Management Sciences, Catonsville, MD.
- Pal, R., and Sinha, K. C. (2002). "Simulation model for evaluating and improving effectiveness of freeway service patrol programs." *J. Transp. Eng.*, 10.1061/(ASCE)0733-947X(2002)128:4(355), 355–365.
- Schrank, D., Lomax, T., and Eisele, B. (2011). "Urban mobility report 2011." Texas Transportation Institute, College Station, TX.
- Sherali, H. D., and Subramanian, S. (1999). "Opportunity cost-based models for traffic incident response problems." *J. Transp. Eng.*, 10.1061/(ASCE)0733-947X(1999)125:3(176), 176–185.
- Skabardonis, A., Petty, K., Bertini, R., Varaiya, P., Noeimi, H., and Rydzewski, D. (1997). "I-880 field experiment: Analysis of incident data." *J. Transp. Res. Board*, 1603(1), 72–79.
- Skabardonis, A., Petty, K., Varaiya, P., and Bertini, R. (1998). "Evaluation of the freeway service patrol (FSP) in Los Angeles." *California PATH Research Rep., Rep. UCB-ITS-PRR-98-13*, Univ. of California, Berkeley, CA.
- Smith, K., and Smith, B. (2001). "Forecasting the clearance time of freeway accidents." *Research Rep. STL-2001-01*, Center for Transportation Studies, Univ. of Virginia, Charlottesville, VA.
- Transportation Research Board (TRB). (1994). "Highway capacity manual: Special report 209." National Research Council, Washington, DC.
- Zhang, H., and Khattak, A. (2010). "What is the role of multiple secondary incidents in traffic operations?" *J. Transp. Eng.*, 10.1061/(ASCE)TE.1943-5436.0000164, 986–997.
- Zhu, S., Kim, W., and Chang, G.-L. (2012). "Design and benefit-cost analysis of deploying freeway incident response units: A case study for Capital Beltway in Maryland." *J. Transp. Res. Record*, 2278, 104–114.