Identifying Secondary Crashes on Freeways Using Sensor Data

Hong Yang, Ph.D. (corresponding author)
Post-Doctoral Associate
Rutgers Intelligent Transportation Systems (RITS) Laboratory
Department of Civil and Environmental Engineering
Rutgers University
623 Bowser Road, Piscataway, NJ 08854, USA
Phone: (732) 445-0579  x127
Fax: (732) 445-0577
Email: yanghong@rci.rutgers.edu

Bekir Bartin, Ph.D.
Research Associate
Rutgers Intelligent Transportation Systems (RITS) Laboratory
Department of Civil and Environmental Engineering
Rutgers University
623 Bowser Road, Piscataway, NJ 08854, USA
Phone: (732) 445-5496
Fax: (732) 445-0577
Email: bbartin@rci.rutgers.edu

Kaan Ozbay, Ph.D.
Professor & Director
Rutgers Intelligent Transportation Systems (RITS) Laboratory
Department of Civil and Environmental Engineering
Rutgers University
623 Bowser Road, Piscataway, NJ 08854, USA
Phone: (732) 445-2792
Fax: (732) 445-0577
Email: kaan@rci.rutgers.edu

Word count: 5471 Texts + 8 Figures = 7471
Abstract: 224
Submission Date: August 1, 2012
Resubmission Date: November 15, 2012

Paper submitted for Publication in the
Transportation Research Record, Journal of Transportation Research Board after being presented
Transportation Research Board’s 92nd Annual Meeting, Washington, D.C., 2013
ABSTRACT
Non-recurring traffic incidents such as motor vehicle crashes increase not only travel delays but also the risk of secondary crashes. Secondary crashes can cause additional traffic delays, and reduce safety. In order to implement effective countermeasures to prevent and/or reduce secondary crashes, first their characteristics should be investigated. However, the related research has been limited largely due to the lack of detailed incident and traffic data necessary to first identify the secondary crashes. Existing approaches such as static methods employed to identify secondary crashes cannot fully capture potential secondary crashes due to fixed spatio-temporal identification criteria. Improved approaches are needed to accurately categorize secondary crashes for further analysis. Therefore, this paper attempts to develop an enhanced approach for identifying secondary crashes using the existing crash database and archived traffic data from highway sensors. The proposed method is threefold: (a) defining secondary crashes; (b) examining the impact range of primary crashes that possibly relate to secondary crashes; and (c) identifying secondary crashes. The proposed methodology establishes a practical framework for mining secondary crashes from existing sensor data and crash records. A case study is performed on a 27-mile segment of a major highway in New Jersey to illustrate the performance of the proposed approach. The results show that the proposed method provides a more reliable and efficient categorization of secondary crashes than commonly used approaches.
INTRODUCTION

Traffic incidents such as motor vehicle crashes often adversely affect normal traffic flow by reducing speed, blocking lanes, inducing queues, and distracting motorists. They account for approximately sixty percent of urban freeway delay (1). Despite their notable impact on mobility, the most serious problem associated with incidents is the risk of secondary crashes (2). For instance, a study by Tedesco et al. (3) found that the crash risk can be six times higher in the presence of a prior crash. Similarly, Karlaftis et al. (4) estimated that each additional minute that a primary incident remains on the roadway increase the likelihood of a secondary crash by about three percent. In addition, studies showed that more than ten percent of crashes were the direct result of previous incidents (5,6,7). In the United States, secondary crashes alone are responsible for an estimated eighteen percent of all fatalities on freeways and twenty percent of all crashes (8,9). The occurrence of secondary crashes can further increase congestion and impede incident management operations. Therefore, it is crucial to examine secondary crashes, understand their characteristics and provide necessary countermeasures.

While traffic incidents have been widely analyzed in the literature, limited research has been conducted to explore secondary crashes. This is largely due to the lack of detailed incident data and corresponding traffic data required to identify secondary crashes (10). It is difficult to identify secondary crashes solely by using the crash databases provided by many transportation agencies. A common method adopted by earlier studies (e.g.,6,7,11,12) is to simply mark the crashes occurred within pre-defined impact range of primary incidents as secondary crashes. However, the impact of a motor vehicle crash on traffic flow depends on various factors such as crash characteristics, roadway conditions, traffic flow levels and weather conditions. Therefore, the validity of secondary crashes identified by this method is questionable.

Traffic sensors such as inductive loop and microwave radar detectors are being instrumented extensively to automate traffic data collection. Large sets of high-quality, high-time resolution traffic data are available for many roadways. A more robust method for identifying secondary crashes thus can be developed by integrating rich traffic data with available crash databases.

Therefore, the objective of this paper is to develop an enhanced method for the identification of secondary crashes that utilizes both archived traffic sensor data and existing crash records.

This paper is organized as follows. The next section provides an in-depth review of existing studies on the identification of secondary crashes. This is followed by the description of the proposed methodology. The proposed methodology is then applied in identifying secondary crashes on a 27-mile section of New Jersey Turnpike. Major findings of our analysis and their implications for future studies are summarized in the final section.

LITERATURE REVIEW

There have been various methods used in the literature to link primary crashes to secondary crashes. The first and the most straightforward method was proposed by Raub (6,7) using static temporal and spatial criteria. It was based on the premise that secondary crashes are those that occur during the clearance period of prior incidents plus post-incident traffic recovery time of 15 minutes, and within a distance of one mile upstream. Based on these criteria, more than 15 percent of crashes on urban arterial roadways were identified as secondary crashes (7). Similarly, Karlaftis et al. (4) adopted these criteria to assess traffic operation on an expressway in Indiana and they found that 35 percent of analyzed crashes could be attributed to primary crashes.

Several other studies also followed the similar concept but slightly adjusted the spatial and/or temporal threshold. For instance, in a study that evaluates a freeway service patrol program in Indiana, Latoski et al. (11) extended the spatial threshold to three miles upstream of a prior crash. Khattak et al. (13) defined secondary crashes by limiting time to the actual duration of primary incidents and space to one mile upstream in the same direction. If an earlier incident blocked lanes, additional 15 minutes were added to the time threshold. Zhan et al. (10) assumed the same temporal criterion as in Raub (6,7), but extended the spatial threshold to two miles upstream of primary incidents in the same direction of travel. In addition, they assumed that only incidents with lane blockage can potentially cause secondary crashes.
In contrast, Hirunyanitiwattana and Mattingly (14) not only established spatial boundary threshold of two miles, but also a temporal threshold of 2 hours to classify primary-secondary crash pair for California highway system. Despite differences in selected thresholds, these aforementioned studies excluded potential secondary crashes in the opposite direction of a primary incident (e.g., rubbernecking crashes).

To address the rubbernecking effect, several studies enhanced the criteria to identify secondary crashes in the opposite direction of primary crashes. For instance, Moore et al. (12) assumed that secondary crashes could occur within two miles upstream in both directions and two hours after an initial incident. They found each primary crash was associated with 0.015 to 0.030 secondary crashes on Los Angeles freeways. The number of secondary crashes was less frequent than the findings in previous studies (4,6,7). The same criteria were used by Kopitch and Saphores (15) on a section of the Interstate 5 in San Diego County and they found that the secondary crashes were 5.53 percent of all crashes. Green et al. (16) determined secondary crashes in Kentucky’s crash database based on the time and distance thresholds of 80 minutes and 6,000 feet, respectively. They also indicated that some secondary crashes do not occur on the same route as the initial incident. Then the length threshold was decreased to 1,000 feet to capture these crashes occurred intersections or side streets. Based on these criteria, the identified secondary crashes only accounted for 0.10 to 0.15 percent of the total annual crashes. Unlike most studies, Chang and Rochon (17) described secondary crashes as the ones that occur within two hours after the occurrence a primary incident, and two miles downstream of the initial one. In addition, the crashes that occur in the opposite direction within 30 minutes from the start of an initial incident and located within a half-mile either downstream or upstream of the initial one were also identified as secondary crashes.

The general method applied in the aforementioned studies is a static approach which used pre-specified temporal and spatial thresholds regardless site and event specific characteristics. The accuracy of a static approach is questionable since it can lead to misclassifications including: (a) overestimation -- the spatio-temporal thresholds may be too large; or (b) underestimation -- the spatio-temporal thresholds may be too small. All in all, there are no fixed spatio-temporal thresholds that can reflect the actual impact of different incidents with varying characteristics.

To address the possible problems when using static thresholds, several queuing-model-based (QMB) approaches were developed using the characteristics of initial incidents. Zhan et al. (18) developed a method to identify secondary crashes based on estimated maximum queue length using cumulative arrival and departure curve technique. Only 3.23 percent of all crashes were classified as secondary crashes as they were within the limits of the estimated maximum queue length and dissipation time. In their previous study, however, they found that the 5.22 percent of all crashes were secondary (10). Sun and Chilukuri et al. (19, 20) proposed a third order polynomial equation to describe the queuing progression curve of initial (primary) incidents. It was found that the secondary freeway incidents identified by static and dynamic thresholds can differ by over 30 percent (21). Zhang and Khattak (22) used a deterministic queuing model to calculate the queue length of each incident to identify secondary crashes in the same direction. Based on the queuing analysis, approximately three percent of all crashes on major freeways in Hampton Roads, Virginia were identified as secondary (22,23,24). Zhang and Khattak (25) pointed out that their earlier method was unable to identify secondary crashes that may have occurred during dissipation of downstream queues. Other studies (26,27) used simulation-based approaches to mimic the impact area of incidents for filtering the secondary ones.

Despite the aforementioned QMB approaches, there is still no universal approach to identify secondary incidents in the literature. The QMB approaches described above improved the quality of identifying secondary crashes. Unlike the static approach, the QMB approaches attempt to construe the prevailing traffic conditions when the primary incident occurs. Several concerns are still associated with the QMB approaches. For instance, the QMB approaches either assumed the impact area of a primary incident reached the maximum length at the clearance time of the primary incident (20) or the spatial impact only existed within the duration of the primary incident (22). These assumptions are debatable because the (maximum) queue length may exist beyond the actual duration of the primary incident (25).

The arrival traffic volume may be sufficiently large that the queue forms faster upstream than it dissipates
downstream, see (28) for an example. Besides, most QMB approaches only accounted for major and obvious primary-secondary incidents such as those occurred in the same direction and/or with lane-blockage (18,19,26,27,29). Some other incidents due to rubbernecking effects or having no lane closure were neglected. Moreover, part of the challenge stems from the lack of accurate information to model or identify the queue formation (26,27,29). Queue lengths based on theoretical queuing models are frequently overestimated (22). There are several studies that used the observed queue information to determine the impact range of primary incidents. For example, Vlahogianni et al. (29) detected incidents and measured vehicle queue lengths using 220 roadside cameras along a 65.2-km urban tollway in Greece. Unfortunately, very few agencies have such a dense surveillance system that can monitor actual queue lengths in the presence of incidents (22). In short, unless very detailed information of each incident is known as well as traffic characteristics and other related information, the QMB approaches cannot accurately identify secondary crashes.

PROPOSED METHODOLOGY
Highway agencies have been using inductive-loop detectors since 1960’s for data collection and traffic surveillance. Much is known about their operation, advantages and disadvantages. As the technology has improved, new and more efficient traffic detection and surveillance devices have also emerged (e.g., video image detectors, microwave radar sensors, and passive acoustic sensors). Today many freeway corridors are equipped with these surveillance units that continuously monitor and archive traffic data [e.g., PeMS in California (28)]. In addition, state DOTs keep and maintain a crash database that include detailed data such as time, location, severity, number of vehicles involved, gender, age and vehicle type. The proposed methodology aims to combine these rich data sources and to fully account for the dynamic spatio-temporal impact of an incident under the prevailing circumstances. The proposed methodology consists of following major components:
(1) defining secondary crashes,
(2) detecting the impact range of a primary incident using archived traffic sensor data, and
(3) identifying secondary crashes within the spatio-temporal impact range of the primary incident.

Secondary Crash Definition
Secondary crashes are stochastic events mainly induced by the impact of a prior incident. FIGURE 1 describes the potential occurrence of secondary crashes under different scenarios. The first scenario describes a scenario where a secondary crash B occurs as a result of the queue caused by the prior incident A. Similarly, the second scenario indicates a case where the prior incident A causes a queue and multiple secondary crashes (e.g., B and C) occur as a result. It should be mentioned that these multiple secondary crashes may also be interdependent. For instance, if an upstream secondary crash (e.g., B) occurs after another secondary crash (e.g., A) and the earlier secondary crash (e.g., B) could be the cause of crash C. Scenarios 1 and 2 represent typical cases that a prior incident caused bottleneck and adversely affected the mobility and safety of a roadway. The third scenario in FIGURE 1 describes the prior incident A causing a secondary crash B, although there is no queue formation. This represents a number of minor incidents that do not seriously impede traffic flow but induce crashes. For instance, an off-road crash that does not cause an obvious queue formation may distract some drivers passing the scene and cause a secondary crash. It is assumed that such secondary crashes can only occur close to the onset time of a primary incident and also within a short distance upstream. The last scenario shown in FIGURE 1 depicts a case where the secondary crash B is caused by the rubbernecking effect. This describes the incurred crashes of vehicles in the opposite direction of the prior incident A.
Impact Range of Primary Incidents

The impact range is the vehicle queue caused by a primary incident measured in time and space. As noted previously, the previous studies estimated vehicle queues either by assuming fixed queue length and duration or by relying on simple queuing modeling techniques. Instead, this study takes advantage of archived traffic sensor data and develops a new methodology to accurately capture the impact range of primary incidents. The following steps describe the methodology.

**Step 1- Constructing Speed Contour Plots:** The impact of each incident on traffic mobility can be visualized through a series of speed contour plots (SCP) over time and space. FIGURE 2 (a) shows an example of a SCP of a freeway under normal traffic condition. The x and y axes represent time and the roadway length, respectively. Each cell in FIGURE 2 (a) represents a speed measurement $V(t,s)$ by traffic sensor $s$ at $t^{th}$ time interval, $\forall s = 1, 2, ..., S$ and $\forall t = 1, 2, ..., T$. Traffic speed is in color coded. FIGURE 2 (b) shows a slice of typical SCP when a crash occurs on the same freeway but on another day. It can be seen that a clear congestion pattern is formed soon after crash A.
Step 2: Constructing a representative speed contour plot: In order to determine what speed range constitutes a queue formation, an average baseline speed range needs to be estimated. Historical sensor data for the days when no incident occurred are used to create a representative speed contour plot (RSCP). Specifically, the $p^{th}$ percentile speed $V_{ip}(t,s)$ is used as the representative speed measured at detector $s$ and time period $t$, where $i=1,2,...,7$ analogous to the day of the week from Monday to Sunday. For instance, we can create a RSCP that represents normal Wednesday traffic conditions on a freeway based on data collected every five minutes during all other Wednesdays in a calendar year. Specifically, the representative speed $V_{ip}(t,s)$ of the RSCP is obtained by the following sub-steps:

(a) Create a subset of historical speed measurement $V_{id}(t,s)$ from detector $s$ at each time period $t$ on $d^{th}$ day of each week, where $d=1,2,...,D$; $D$ is the total number of all sampled days that belong to the $i^{th}$ category. For instance, if $i=3$, $D$ represents the total number of Wednesdays during a period (e.g., a calendar year) for which historical data are available;

(b) Sort the historical speed measurements $V_{j}(t,s)$ in ascending order. Denote $V_{j}(t,s)$ as the sorted speed measurement at $j^{th}$ place, $j=1,2,...,D$; and

(c) Let $k$ be an integer such that $k = \text{Int}(p \times D) + 1$, where $\text{Int}(x)$ denotes integral part of a number $x$. Find the $V_{ip}(t,s)$ such that $V_{ip}(t,s) = V_{k}(t,s)$.

If there is recurring traffic congestion at a specific section of the freeway, the corresponding representative speeds shown in RSCP should be low (e.g., red cells in RSCP). In contrast, if there is no recurring congestion, the RSCP should reflect very smooth traffic conditions [e.g., most of measurements are high speed (green cells) in RSCP]. Identifying the recurring congestion is very useful as the incident-induced congestions can be separated from the recurrent congestion. However, it should be mentioned that serious recurrent congestion at some sections can form long queue upstream. Primary-secondary crash pair occurred within the queue cannot be practically determined. This is because that the recurrent queue covered the congestion caused by the prior incident. In such case, it is difficult to clearly judge whether a later crash was caused by the prior incident or caused by the recurrent congestion. Thus, such confounded case was excluded in the following step.
Step 3 - Constructing a binary speed contour plot (BSCP): Compare $V(t,s)$ of SCP in step 1 with the corresponding $V_{w}(t,s)$ in RSCP generated in step 2. If $V(t,s) > \omega \times V_{w}(t,s)$, the speed measurement $V(t,s)$ in the original SCP is converted into $\hat{V}(t,s) = 1$. Otherwise, it is denoted as $\hat{V}(t,s) = 0$. This constraint defines the abnormal traffic condition if the measured speed is below the threshold $\omega \times V_{w}(t,s)$. Here $\omega$ is a user defined weighting factor between 0 and 1. It assumed that a $(1-\omega) \times 100$ percent reduction in the representative (normal) speed indicates the occurrence of congestion. A small $\omega$ suggests an aggressive threshold to define congestion whereas a large $\omega$ implies a conservative threshold. To reflect the consistency of speed measurements in short time periods, the speed measurement at the $t^\text{th}$ time period will be changed to $\hat{V}(t,s) = 1$ if $|\hat{V}(t-1,s) - \hat{V}(t,s)| < |\hat{V}(t-1,s) - \hat{V}(t+1,s)|$. After conversion, the original SCP will be represented by a BSCP. FIGURE 3 shows an example of converting the original SCP into a BSCP based on the RSCP of 50th percentile historical speeds. In BSCP, a red cell indicates that $\hat{V}(t,s) = 1$ and a green cell means $\hat{V}(t,s) = 0$. Each cluster of red cells visualizes non-recurring traffic congestion associated with an incident [e.g., crash A in FIGURE 3 (b)].

![Figure 3](image-url)

**FIGURE 3** Converting SCP into a binary speed contour plot.

### Identifying Secondary Crashes

We can identify the queue caused by incidents using the steps described in the previous section. The next task is to identify whether a crash is associated with a prior incident. Let us recall the BSCP generated in FIGURE 3, and assume that there are two more crashes B and C that happened on the same day after the first crash A. The time and location of each crash are shown in FIGURE 4. The task is to determine whether B and C are secondary crashes related to the first crash A. If we use the static method or queue-based method that assumes the maximum temporal and spatial impact range as thresholds, both crash B and C have to be classified as secondary crashes. However, crash C should not be a secondary crash as the queue triggered by crash A ($t = 08:15, s = S17$) has not reached the location of crash C at the time when it occurred ($t = 08:28, s = S23.5$). In contrast, when crash B occurred ($t = 09:37, s = S21.6$) it was in the queue. Therefore, only crash B should be classified as a secondary crash.
However, it is time consuming to identify secondary crashes visually as it is done above. Therefore, we have developed an algorithm that automatically identifies secondary crashes within the impact range. The algorithm is described as follows:

**Step 1 - Estimate the Equation of a Straight Line between a Pair of Crashes:** Using the coordinates of the prior incident and a potential secondary crash, a line that links the two crashes are formed. Using FIGURE 5 as an example, the line AB and line AC can be represented using the following equations:

Line AB: \[ s_x = \frac{(s_B - s_A)}{(t_B - t_A)} \times (t_x - t_A) + s_A \]

Line AC: \[ s_x = \frac{(s_C - s_A)}{(t_C - t_A)} \times (t_x - t_A) + s_A \]

where \((t_A, s_A)\), \((t_B, s_B)\), and \((t_C, s_C)\) are coordinates (in terms of time and distance) of crash A, crash B, and crash C, respectively. The pair \((t_x, s_x)\) represents the coordinate of any point \(X\) on an estimated line (i.e., line AB).

**Step 2 - Make Two Types of Predictions Using the Estimated Line Equation in Step 1:** Type I -- given \(t_x = t_{a,1}, t_{a,2}, ..., t_{a,n}\), predict \(s_x\), assuming that \(t_m \leq t_r < t_{a,i}\) and \(t_r \leq t_x < t_{a,i+1}\). Type II -- given \(s_x = s_{h,1}, s_{h,2}, ..., s_{h,n}\), predict \(t_x\), assuming that \(s_r \leq s_t < s_{h,i}\) and \(s_t \leq s_{h,i+1}\). FIGURE 6 (a) and FIGURE 6 (b) show the examples of these two types of predictions, respectively.

**Step 3 - Calculate the Coordinates of Middle Points:** The midpoint of any two adjacent points predicted in step 2 is calculated. FIGURE 6 (c) illustrates an example of \(Q\) middle points \(M_1, M_2, ..., M_q, M_{q+1}, ..., M_H\) that have to be calculated. The coordinate of each middle point is denoted by \((t_q, s_q)\).

**Step 4 - Check Binary Speed Measurements \(\hat{V}(t,s)\) of the Cells that Enclose Middle Points:** Let \(t_k < t_x < t_{k+1}\) and \(s_k < s_x < s_{k+1}\), where \(t_k\) is the \(R\)th time interval in BSCP and \(s_k\) is the location of the \(L\)th sensor. We can easily find the cell that contains the \(q\)th middle point based on the information of the coordinates of the point [see FIGURE 6 (d) as an example]. Assume the binary speed measurement of the identified cell is \(\hat{V}_q(t_{x}, s_x)\), where \(\hat{V}_q(t_{x}, s_x) = 1\) indicates congestion and \(\hat{V}_q(t_{x}, s_x) = 0\) means normal condition.
If the binary speed measurements of all corresponding middle points are equal to 1 [i.e., FIGURE 6 (e)], we have \( \sum_{q=1}^{Q} \hat{\nabla} \gamma(t_{r}, s_{r}) = Q \). This condition indicates that the line between the prior incident and a later crash is located in the impact range of the prior incident. Once we confirm that the line is located within the impact range, the later crash is identified as a secondary crash [e.g., crash B in FIGURE 6 (e)]. However, if a part of the line is not located in the impact range of the prior incident, we have \( \sum_{q=1}^{Q} \hat{\nabla} \gamma(t_{r}, s_{r}) < Q \). In other words, the later crash is not affected by the prior one if a portion of the line is out of the impact range. Thus the later crash is not classified as a secondary crash [e.g., crash B in FIGURE 6 (f)].

**FIGURE 6** Visual depiction of the proposed algorithm that determines the relationship between two crashes.
CASE STUDY

The proposed methodology is used to identify secondary crashes on a 27-mile section of a major highway in New Jersey.

Data Description

A section of New Jersey Turnpike located between interchanges 5 and 9 was used as a case study. FIGURE 7 shows the site map. The study section is about 27 miles. It has 25 traffic sensors placed approximately at every one mile of the mainline. There are 3 lanes between milepost 48 and milepost 72.5 and 5 lanes beyond milepost 72.5 towards north, as shown in FIGURE 7. The posted speed limit of the section is 65 mph. Traffic data collected by each traffic sensor in 2011 are used for the analyses. The data include volume, speed, and occupancy measured at different time intervals at each sensor station. The data aggregated in 5-minute intervals were extracted to develop the speed contour plots using the proposed method.

The corresponding crash records in 2011 were obtained from the crash database of the New Jersey Department of Transportation (NJDOT) available online (30). The records include the detailed crash characteristics such as date, time, location, crash type, number vehicles involved, vehicle characteristics, and crash severity. After the exclusion of two crashes occurred on ramps and other 24 with unknown direction, the remaining data consist of a total of 1,188 crash records for the studied section (Southbound: 575 crashes; Northbound: 613 crashes).

Implementation of the Proposed Methodology

To identify the potential secondary crashes on a given day, the corresponding SCP, RSCP, and BSCP need to be generated. The daily SCP is created based on the speed measurements from the traffic sensors. The 50th percentile value of the historical speed data is used to create the RSCP based on the proposed method. It is assumed that the weighting factor $\alpha = 0.7$ when converting the SCP to BSCP. In other words, a 30 percent reduction in the representative (normal) speed is identified as traffic congestion. This weighting factor is determined based on previous studies related to congestions and/or bottlenecks analysis on freeways (e.g., 31,32,33,34). Based on the SCP, RSCP, and the selected weight factor, BSCP for each day is generated. The observed crashes are also superimposed on the BSCP. Then the BSCP together with the proposed identification algorithm are then used to identify the major types of secondary crashes shown in FIGURE 1 (a) and FIGURE 1 (b). The third type of secondary crash shown in FIGURE 1 (c) was assumed to be crashes occurred within 30 minutes and 0.5 miles upstream of the prior incident. The forth type of secondary crashes is defined as the crashes that occurred in the opposite direction within one hour and one mile upstream of the prior incident. In addition to the spatial and temporal criteria, the potential secondary crash in the opposite direction has to be in a queue to be considered as a secondary crash due to rubbernecking effect.
For comparison purposes, secondary crashes are also identified using the static method. Sixteen types of fixed spatio-temporal boundaries ranging from 0.5 to 2 miles and from 0.5 to 2.0 hours were identified as secondary crashes in the same direction of a prior incident. For possible secondary crashes in the opposite direction of a prior incident, the same criteria described in previous paragraph were used.

Both the proposed method and the static approach were coded and implemented in the statistical software package, R. The results are summarized in the following section.

**Results and Discussion**

The secondary crashes are identified using both the static and the proposed methods. The static method identified 42 out of 1,188 crashes (3.5 percent) as secondary crashes based on the maximum spatio-temporal boundaries of two miles and two hours. If the spatio-temporal thresholds of 0.5 miles and 0.5 hours are used the static method identifies only 18 secondary crashes. If the static thresholds were increased, the static method can capture more secondary crashes. However, the false identifications will also increase. As mentioned earlier, selection of spatio-temporal criteria is subjective and the static method cannot account for different types of incidents that have varying impact range (in time and space).

For instance, FIGURE 8 (b) shows a crash (time= 17:30) occurred within two miles upstream of the prior crash (time=15:37). If the static method were used, the crash have been classified as a secondary crash. However, the original speed contour plot in FIGURE 8 (a) clearly suggests that the prior crash (time=15:37) had no notable impact on overall traffic flow. In addition, it was not the scenario described in FIGURE 1 (c) where the later crash is in the vicinity of a prior crash. Thus the later crash (time= 17:30) was unlikely to be caused by the prior incident.

In contrast, the proposed method individually examines the impact of each incident and identified 100 secondary crashes as a result of 71 primary crashes. For a sensitive analysis, if we use an aggressive weighting factor $\omega=0.65$, 80 secondary crashes were identified. In contrast, if $\omega$ was increased to 0.75, 116 secondary crashes were identified. The variations of the results suggest that a reasonable weight factor should be used based on the highway operators' definitions of normal and congested traffic. Despite the variation, all the numbers are twice or more than the number (42 crashes) identified by the static method. The large difference between the findings of the two methods was expected because the static method cannot capture many primary incidents that impact traffic beyond the defined boundaries. For instance, a major crash such as a multi-vehicle collision blocking multiple lanes can create multiple hours and/or miles of disruption of traffic (e.g., 6,7,28). Thus, many secondary crashes that occurred beyond the defined boundaries were not identified by the static method but by the proposed method, see FIGURE 8 (b) for an example.

23 of the secondary crashes identified by the static method were also classified as secondary crashes by the proposed method. The remaining 19 [like the one identified by the static method in FIGURE 8 (b)] were successfully excluded by the proposed method. FIGURE 8 (c) and FIGURE 8 (d) illustrate the case when both methods identified the same secondary crashes. In this case, the secondary crash was not only within the pre-defined spatio-temporal boundaries of a prior crash but also in the queue triggered by it. Other than the single primary-secondary crash pair, the proposed method can also determine multiple secondary crashes in relation to a primary one. For instance, FIGURE 8 (e) and FIGURE 8 (f) illustrated that a crash (at 14:57) occurred in northbound of the studied freeway induced two secondary crashes (at 15:58 and 16:26) at the end of the long queue, which was far beyond the spatial boundary of the static method.

In summary, the static method only identified 42 secondary crashes whereas the proposed method identify 100 secondary crashes based on the assumed weighting factor. These results indicate that the proposed method not only reduces the incorrect classifications but also captures more number of secondary crashes missed by the static method. Another advantage of the proposed method is its computational efficiency. Our case study showed that it only took two to three minutes to process 24-hour archived sensor data and complete the identification.
FIGURE 8 Identified secondary crashes using different methods.
CONCLUSIONS

This study extended the previous research on identifying secondary crashes. A new methodology that takes advantage of archived traffic data from traffic sensors was presented. Instead of assuming fixed spatio-temporal thresholds, the proposed methodology based on the concept of the Binary Speed Contour Plot (BSCP) captures the evolution of actual traffic queue induced by a primary incident. Based on the detected impact range of each primary crash, the proposed identification method automatically examines potential secondary crashes caused by the prior incident. It should be mentioned that the proposed framework also integrated the static method to cover cases when no queue exists. Step-by-step description of the proposed approach provides a readily deployable data analysis tool for transportation agencies to identify secondary crashes through archived sensor data and existing crash databases.

A case study to test the performance of the proposed methodology was also presented. The secondary crashes induced by primary crashes were identified. The results showed that the proposed methodology can identify twice the number of secondary crashes than that of the static method based on fixed spatio-temporal thresholds. This was because the proposed methodology successfully captured crashes with impact far beyond the pre-defined spatio-temporal thresholds by the static method. In addition, 45 percent of secondary crashes identified by the static method were excluded by the proposed method as there was weak evidence to link them with prior incidents. Although the case study demonstrated the secondary crashes identified by primary crashes, the same approach can be used to examine secondary crashes caused by other non-crash incidents given the availability of these incident data. The proposed methodology reduced the biases (overestimation and underestimation) associated with the subjective thresholds used by the static method. In addition to the superior performance of the proposed identification approach, it only took a short time period to aggregate the daily sensor data and crash information to identify whether a primary incident caused secondary crashes. With the availability of real-time sensor data, this kind of computational efficiency enables the proposed method to conduct faster identification of secondary crashes. Given the access to real-time sensor data, the proposed methodology can be easily implemented to conduct on-line identification of secondary crashes. More studies with high-quality sensor data to investigate the capability of the proposed approach are suggested.

ACKNOWLEDGEMENTS

The contents of this paper reflect views of the authors who are responsible for the facts and accuracy of the data presented herein. The contents of the paper do not necessarily reflect the official views or policies of the agencies. Thanks to seven anonymous reviewers for their helpful comments and suggestions to improve this paper.

REFERENCES


24. Khattak, A., X. Wang, and H. Zhang. Spatial Analysis and Modeling of Traffic Incidents for Proactive Incident Management and Strategic Planning. In Transportation Research Record:


