# Identification and Evaluation of Secondary Crashes on Wisconsin State Trunk Highways

Dongxi Zheng, M.S.
Graduate Research Assistant

Madhav Chitturi, Ph.D.
Associate Researcher

Andrea Bill, M.S.
Traffic Safety Engineering Research Program Manager

David A. Noyce, Ph.D., PE
Professor and Director, TOPS Lab

WISCONSIN TOPS
TRAFFIC OPERATIONS & SAFETY LABORATORY

March 2014
Submitted to Wisconsin Department of Transportation

By

Wisconsin Traffic Operations and Safety (TOPS) Laboratory
University of Wisconsin-Madison
Department of Civil and Environmental Engineering

## DISCLAIMER

This research was funded by the Wisconsin Department of Transportation. The contents of this report reflect the views of the authors who are responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Wisconsin Department of Transportation, Federal Highway Administration or the University of Wisconsin.

This document is disseminated under the sponsorship of Department of Transportation in the interest of information exchange. The agencies listed above assume no liability for its contents or use thereof. This report does not constitute a standard, specification, or regulation, and its contents are not intended for construction, bidding, or permit purposes.

The name of any products or manufacturers listed herein does not imply an endorsement of those products or manufacturers. Trade and manufacturers' names appear in this report only because they are considered essential to the object of the document.

| | | |
|---|---|---|
| 1. Report No. | 2. Government Accession No. | 3. Recipient's Catalog No. |
| 4. Title and Subtitle<br>Identification and Evaluation of Secondary Crashes on Wisconsin State Trunk Highways | | 5. Report Date<br>March 2014<br>6. Performing Organization Code |
| 7. Author(s) Dongxi Zheng, Madhav Chitturi, Andrea Bill, and David A. Noyce | | 8. Performing Organization Report No. |
| 9. Performing Organization Name and Address<br>Traffic Operations and Safety Laboratory<br>1415 Engineering Dr. Room B239<br>Madison, WI 53706 | | 10. Work Unit No. (TRAIS)<br>11. Contract or Grant No. |
| 12. Sponsoring Agency Name and Address<br>Wisconsin Department of Transportation | | 13. Type of Report and Period Covered<br>14. Sponsoring Agency Code |
| 15. Supplementary Notes | | |

16. Abstract

The annual cost of congestion in the United States reportedly exceeds $120 billion. Freeway incidents are major sources of non-recurrent congestion and the resulting secondary crashes can prolong the traffic impact and increase the cost. Research on secondary crashes to support statewide transportation system management has been limited. In the current study, an efficient and effective procedure was developed for identifying secondary crashes from statewide historical data. This procedure is a two-part process. In the first part, a three-stage filtering method is performed to exclude irrelevant input crashes and a list of crash pairs that are likely to be primary-secondary crash pairs is generated. In the second part, researchers manually validate actual secondary crashes from the pairs by reading through the crash reports. Efficiency is gained through the first part of the procedure, which is fully automated in a custom java program. Utilizing a linear referencing system for crash pairing in the program is a major advantage over a 2-dimensional spatial search, making the program much faster and more accurate than an ArcGIS based counterpart.

A total of 302 secondary crashes were identified over a five-year time span on about 1,500 miles of Wisconsin access controlled highways. Rear-end crashes compose 76% of the 302 secondary crashes. Secondary crashes happening in the same traffic direction of the primary crash occurred about twice as often as those in the opposite direction. Secondary crashes and general crashes have similar trends related to the day of week. However, more secondary crashes (by percentage) than general crashes happened in late night hours and November-January; fewer secondary crashes than general crashes happened in May-June. Secondary crash hotspots are normally within 1-mile from an interchange. Inattentive driving conditions, debris, construction zones, and visibility obstructions (e.g., fog) are leading highway factors to secondary crashes. Following too closely, inattentive driving, failure to control vehicle, and speeding are leading factors on secondary crashes from the driver side. The average time lapse between a primary crash and a secondary crash is 17 minutes. The average distances from an upstream secondary crash to the primary crash are 0.29 miles and 0.37 miles for the same side of traffic and the opposite side of traffic, respectively; the average distances from a downstream secondary crash to the primary crash are around 0.15 miles for both traffic directions. Congestion could cause secondary crashes in both upstream and downstream traffics in both traveling directions. Police officers are recommended to increase their investigation scope and attention for secondary crashes when traffic is congested.

| | | | |
|---|---|---|---|
| 17. Key Word<br>Secondary Crash, Traffic Incident Management, Large Scale Network | | 18. Distribution Statement | |
| 19. Security Classif. (of this report) | 20. Security Classif. (of this page) | 21. No. of Pages<br>61 | 22. Price |

**Form DOT F 1700.7** (8-72) Reproduction of completed page authorized

## ABSTRACT

The annual cost of congestion in the United States reportedly exceeds $120 billion *(1)*. Freeway incidents are major sources of non-recurrent congestion and the resulting secondary crashes can prolong the traffic impact and increase the cost. Research on secondary crashes to support statewide transportation system management has been limited. In the current study, an efficient and effective procedure was developed for identifying secondary crashes from statewide historical data. This procedure is a two-part process. In the first part, a three-stage filtering method is performed to generate a list of potential primary-secondary crash pairs. In the second part, researchers manually validate actual secondary crashes from the pairs by reading through the crash reports. Efficiency is gained through the first part of the procedure, which is fully automated in a custom program written in Java™. Utilizing a linear referencing system for crash pairing in the program is a major advantage over a 2-dimensional spatial search, making the program much faster and more accurate than an ArcGIS based counterpart.

A total of 302 secondary crashes were identified over a five-year time span on about 1,500 miles of Wisconsin access controlled highways. Rear-end crashes compose 76% of the 302 secondary crashes. Secondary crashes happening in the same traffic direction as the primary crash occurred about twice as often as those in the opposite direction. Secondary crashes and general crashes have similar trends related to the day of week. However, more secondary crashes (by percentage) than general crashes happened in late night hours and November-January; fewer secondary crashes than general crashes happened in May-June. Secondary crash hotspots are normally within 1-mile from an interchange. Inattentive driving conditions, debris, construction zones, and visibility obstructions (e.g., fog) are leading highway factors for secondary crashes. Following too closely, inattentive driving, failure to control vehicle, and speeding are leading driver factors for secondary crashes. The average time lapse between a primary crash and a secondary crash is 17 minutes. The average distances from an upstream secondary crash to the primary crash are 0.29 miles and 0.37 miles for the same side of traffic and the opposite side of traffic, respectively; the average distances from a downstream secondary crash to the primary crash are around 0.15 miles for both traffic directions. Congestion could cause secondary crashes in both upstream and downstream traffics in both traveling directions. Police officers are recommended to increase their investigation scope and attention for secondary crashes when traffic is congested.

# TABLE OF CONTENTS

**DISCLAIMER** ..... **i**

**Abstract** ..... **iii**

**Table of Contents** ..... **iv**

**List of Figures** ..... **vi**

**List of Tables** ..... **viii**

**Chapter 1 Introduction** ..... **9**

1.1 Problem Definition ..... 9

1.2 Study Objectives ..... 10

1.3 Research Scope ..... 10

1.4 Report Organization ..... 10

**Chapter 2 Literature Review** ..... **12**

2.1 Secondary Crash Studies ..... 12

2.1.1 Definition and Identification of Secondary Crashes ..... 12

2.1.2 Influential Factors on Secondary Crashes ..... 18

2.2 Traffic Incident Management ..... 21

2.2.1 Who and What Are Involved in Traffic Incident Management ..... 21

2.2.2 How to Evaluate the Performance of Traffic Incident Management ..... 23

2.2.3 How to Improve Traffic Incident Management System ..... 23

2.2.4 Related Works ..... 24

2.3 Freeway Service Patrols ..... 25

2.4 Summary ..... 26

**Chapter 3 Secondary Crash Identification Method** ..... **27**

3.1 Raw Data Description ..... 28

3.1.1 STN Links and Linear Referencing ..... 29

3.1.2 Crash Records ..... 30

3.1.3 Other Data ..... 30

3.2 Stage 1: Initial Filtering ..... 31

3.2.1 On selected STN highways ..... 31

3.2.2 Non-deer related ..... 31

3.2.3 Non-inclement weather related ....................31

3.3 Stage 2: Time-Space Pairing ....................32

3.3.1 Local Linear Coordinate System ....................33

3.3.2 Candidate Link ....................35

3.3.3 The Algorithm ....................36

3.3.4 Extension to Data Evolution ....................37

3.3.5 Validation ....................38

3.4 Stage 3: Pair-wise Filtering ....................39

3.4.1 Ramp Filter ....................39

3.4.2 Impact Area Filter ....................40

3.5 Review of Crash Reports ....................42

**Chapter 4 Secondary Crash Identification Results ....................44**

**Chapter 5 Analysis ....................46**

5.1 Temporal Trends of Secondary Crashes ....................46

5.2 Secondary Crash Hotspots ....................48

5.3 Contributing Factors to Secondary Crashes ....................54

5.4 Spatial and Temporal Relationship between Primary Crashes and Secondary Crashes 55

**Chapter 6 Conclusions ....................57**

**References ....................59**

## LIST OF FIGURES

Figure 1 Spatial study scope. ........ 11
Figure 2 Evolution of secondary crash identification methods. ........ 13
Figure 3 Summary of influential factors on secondary crash occurrence in the literature. ........ 18
Figure 4 An automatic method to find secondary crash candidates. ........ 28
Figure 5 Illustration of STN chains, links, and crashes in ArcGis. ........ 29
Figure 6 Illustration of the linear referencing system. ........ 30
Figure 7 Four possible distances between two crashes based on the relationship vector. ........ 35
Figure 8 Pseudo code of crash pairing algorithm. ........ 36
Figure 9 User interface of the crash pairing program. ........ 38
Figure 10 Sampling intervals. ........ 40
Figure 11 Illustration of an impact area. ........ 42
Figure 12 Illustration of the manual review environment. ........ 43
Figure 13 Filtering results of primary-secondary candidate pairs. ........ 44
Figure 14 Crash distribution by month. ........ 47
Figure 15 Crash distribution by day of week. ........ 47
Figure 16 Crash distribution by hour of day. ........ 47
Figure 17 Histogram of secondary crash frequencies (2007 – 2011). ........ 48
Figure 18 Statewide overview of secondary crash frequencies. ........ 49
Figure 19 Secondary crash frequency in Milwaukee (2007-2011). ........ 50
Figure 20 Secondary crash frequency in Madison (2007-2011). ........ 51
Figure 21 Secondary crash frequency in Green Bay (2007-2011). ........ 52
Figure 22 Secondary crash frequency in Appleton (2007-2011). ........ 53
Figure 23 Secondary crash frequency by highway factor. ........ 54
Figure 24 Secondary crash frequency by driver factor. ........ 54
Figure 25 Time distance distribution. ........ 55

Figure 26 Network distance characteristics. ........................................................................ 56

## LIST OF TABLES

Table 1 TIM Related Agencies and Responsibilities.................................................................... 22

Table 2 2009 TIMSA Questionnaire Overview............................................................................. 23

Table 3 Manual Review Results ..................................................................................................... 45

Table 4 Overview of Identified Secondary Crashes ...................................................................... 46

# CHAPTER 1 INTRODUCTION

The annual cost of congestion in the United States reportedly exceeds $120 billion *(1)*. Freeway incidents are major sources of non-recurrent congestion and the resulting secondary crashes can prolong the traffic impact and increase the cost. A secondary crash is an undesirable consequence resulting from a primary incident. More formally, according to the Federal Highway Administration (FHWA), "secondary crashes are those that occur with the time of detection of the primary incident where a collision occurs either a) within the incident scene or b) within the queue, including the opposite direction, resulting from the original incident" *(2)*. Existing studies have shown the extended traffic impact and the economic costs of secondary crashes *(3–5)*. Therefore, reducing the chances of secondary crashes becomes a major consideration in the dispatch plans of traffic incident management (TIM) agencies *(6, 7)*.

In spite of various findings on secondary crashes, most existing studies were limited by scope. Many studies were conducted on only one or two sample freeways or a short segment of highway; other studies extended the scope to freeways but considered a small regional scale. Only two studies were performed on a large scale that involved statewide highway systems. One of the major reasons for such scope constraints was the challenge of secondary crash identification. In order to identify secondary crashes accurately, most existing studies considered the dynamic features of the traffic impact caused by the primary incidents. Thus, the study scopes were limited to highway facilities where high resolution traffic data were available for dynamic analyses. In addition, modeling the dynamic impact of primary incidents required considerable computational efforts, which could be too heavy or even infeasible for a statewide transportation system. Previous studies considering statewide highway systems did not consider the dynamic impact (i.e., time-varying queue lengths) caused by primary incidents. In summary, none of the previous studies investigated secondary crashes on a statewide transportation system while considering the dynamic impact of primary incidents.

To fill the above research gap, the current study developed a three-stage method to automatically narrow down the search space for secondary crashes from massive historical data. Due to the utilization of linear referencing, the three-stage method is efficient for large scale networks such as a state access controlled highway network. Secondary crash candidates generated from the three-stage method are validated via manual review of crash reports. Analyses were done on the identified secondary crashes.

## 1.1 Problem Definition

The research team is not aware of any existing statewide secondary crash study for Wisconsin, even though there is a promising database for such purposes. The Wisconsin Department of Transportation (WisDOT) maintains statewide traffic, highway, and crash data, through the fusion of which, secondary crashes can be identified and studied. Challenges come from the size and the complexity of the database, which invalidate most existing secondary crash identification methods in terms of their efficiency or accuracy. In addition, the characteristics of secondary crashes are unknown. Finding these characteristics and using them in the context of traffic incident management are important and beneficial.

## 1.2 Study Objectives

In general, the study objectives were to 1) establish a repeatable methodology that allows practitioners to collect and analyze secondary crashes based on historical data; and 2) provide any relevant findings about secondary crashes as references for TIM decision makers. The following are specific objectives:

- Conduct a comprehensive literature review about secondary crashes, traffic incident management, and freeway service patrols;
- Give a comprehensive definition of secondary crashes;
- Develop an efficient and effective method to identify secondary crashes from a statewide highway network; and
- Validate and analyze the identified secondary crashes in terms of their temporal trends, dispersion, influential factors, and relationship to primary crashes.

## 1.3 Research Scope

This research is limited to secondary crashes that occurred on approximately 1,500 miles of selected access controlled highways per direction of travel in Wisconsin between 2007 and 2011. The selected highways are illustrated in Figure 1. These highways include US highways, interstate highways, and state trunk highways.

## 1.4 Report Organization

Aside from this introduction, the rest of this report is organized into five other chapters. Chapter 2 presents a thorough literature review on secondary crash studies, traffic incident management, and freeway service patrols to identify research gaps. Chapter 3 develops the procedure to identify secondary crashes from historical data. Chapter 4 shows the secondary crash identification results on the selected highway network over a five year time span. Chapter 5 presents temporal and spatial analyses on the identified secondary crashes, influential factors of secondary crashes, and examines the characteristics of the temporal and spatial displacements of secondary crashes from primary crashes. Chapter 6 concludes the report by highlighting the findings and making recommendations for future research.

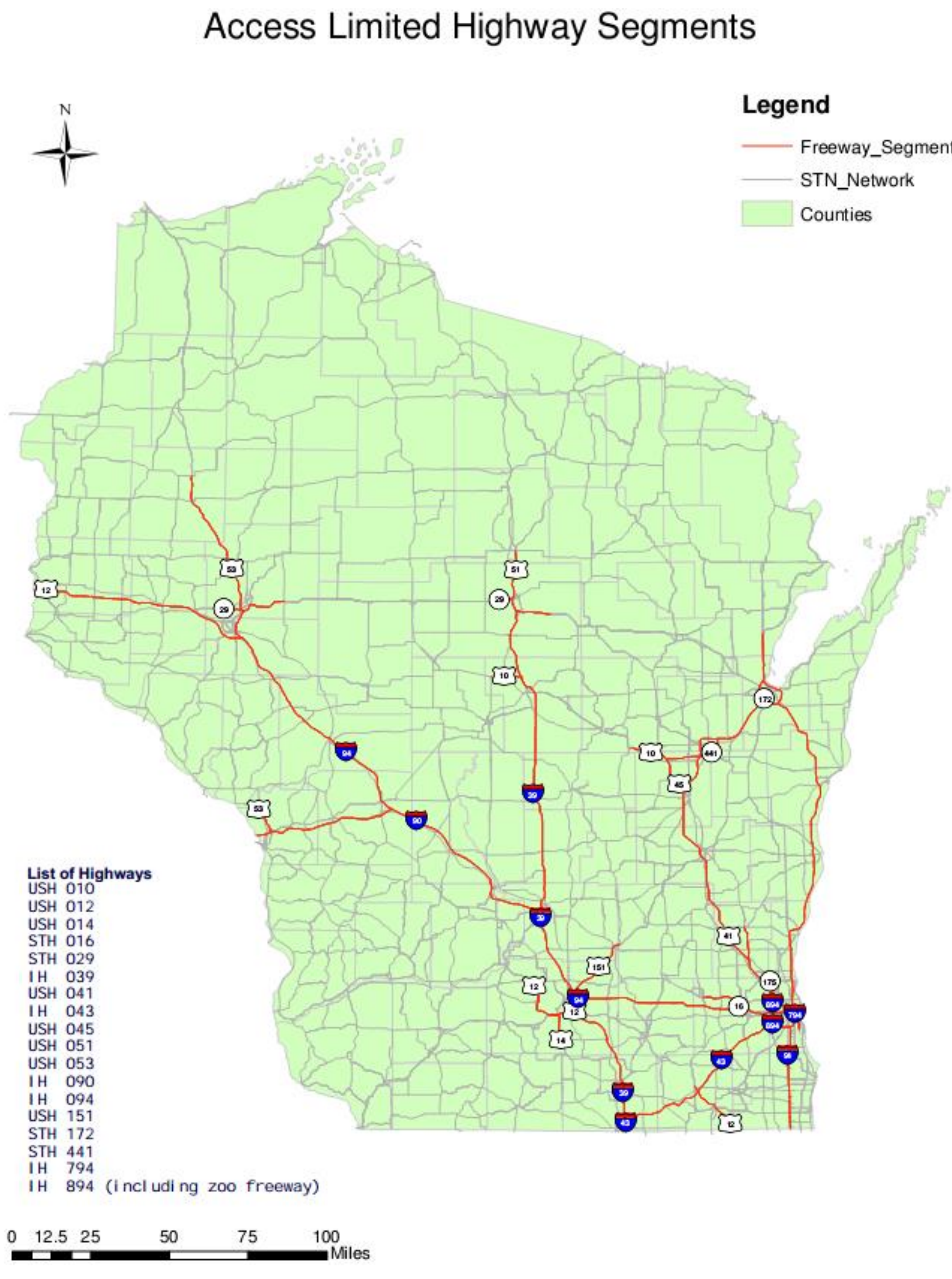



**FIGURE 1 Spatial study scope.**

# CHAPTER 2 LITERATURE REVIEW

In this chapter, a literature review is presented on the following three topics: 1) secondary crash identification and causal analysis; 2) evaluation of traffic incident management programs; and 3) freeway service patrols. The primary focus is on identification of secondary crashes, which reveals the research gap of missing an efficient algorithm for secondary crash data mining in a large scale highway network. Reviews of influential factors on secondary crashes, traffic incident management programs, and freeway service patrols provide the research team with a better idea of what can be analyzed with respect to identified secondary crashes so that practical recommendations can be made to the Wisconsin Department of Transportation, as well as national transportation practitioners.

## 2.1 Secondary Crash Studies

Secondary crashes have been studied since the 1970s. Two major research directions for secondary crashes are documented in the literature: 1) defining and identifying secondary crashes from existing databases and 2) discovering influential factors and modeling the likelihood of secondary crash occurrences. With improved incident reporting techniques and the emergence of intelligent transportation systems (ITS), researchers have expanded on both the definition and the identification procedures used in order to capture secondary crashes more accurately. Nevertheless, a uniform definition of secondary crashes has not been reached. Additionally, even with the same definition, different identification procedures can be used according to the data variety; the accuracy of results is further constrained by data quality. Previous researchers have proposed different influential factors based on secondary crashes identified by different methods. Some researchers proposed logistic models to estimate the likelihood of secondary crash occurrence using the influential factors found. A recent practice by Green et al. contained a good summary of some relevant studies on secondary crashes *(8)*. To improve the completeness and timeliness of this literature review, additional studies other than those mentioned in Green et al *(8)* are also included. The rest of the literature review is divided into two sections: the first section is focused on secondary crash definition and identification and the second section reviews the analysis of secondary crash characteristics.

### 2.1.1 Definition and Identification of Secondary Crashes

According to the Federal Highway Administration, "secondary crashes are those that occur with the time of detection of the primary incident where a collision occurs either a) within the incident scene or b) within the queue, including the opposite direction, resulting from the original incident" *(2)*. The identification of secondary crashes has been a research challenge for recent decades. Data availability and quality are major issues. For example, to identify the queue length caused by the primary incident, traffic data are needed. However, even with the recent emergence of intelligent transportation systems (ITS), data accuracy is still limited such that traffic conditions can only be observed or derived up to a certain confidence level. In addition to the difficulty of identifying a secondary crash, verifying the causal relationship between the primary crash and the secondary crash is not trivial. Verifications are normally done through reviewing police incident reports. Keywords such as "another crash" in the police investigations are typically searched to confirm the connection between two incidents. Unfortunately, such clues are not always available. For example, a secondary crash occurring at the end of a long queue might not be recognized as being related to the primary incident due to limited on-site visibility and communications when the police officers were investigating it. The rest of this section is focused on the evolution of secondary crash identification methods. Each of the methods was generally coupled with the corresponding researcher's definition of secondary crashes. Figure 2 is an illustration of the three major families of identification methods to be reviewed.

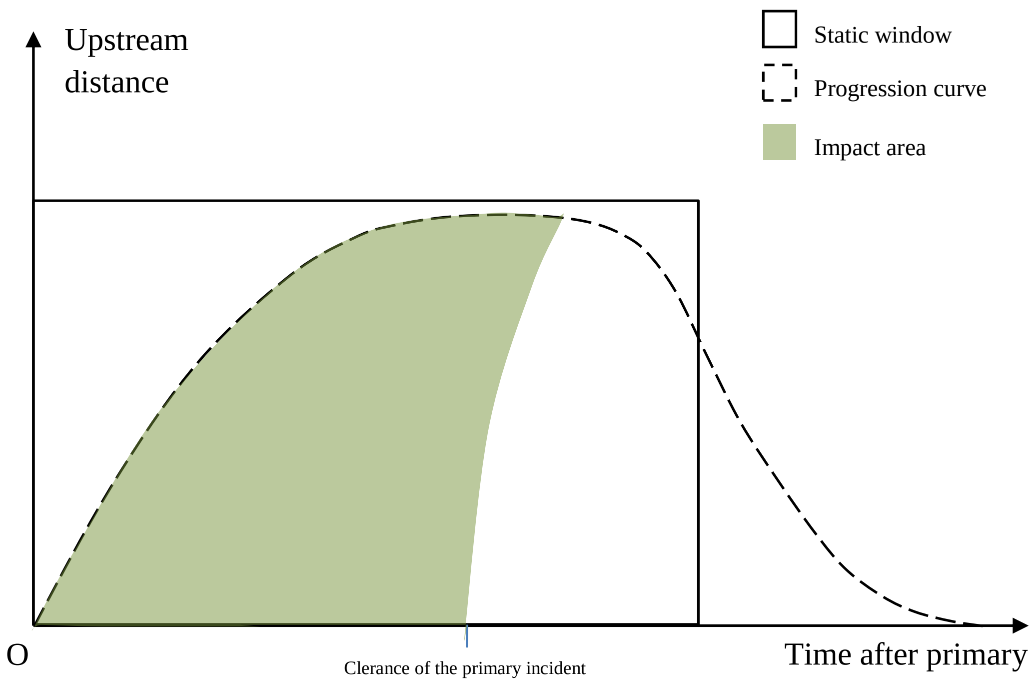



**FIGURE 2 EVOLUTION OF SECONDARY CRASH IDENTIFICATION METHODS.**

***Static spatial and temporal thresholds***

A simple, intuitive, and widely used method to identify secondary crashes is by defining them as crashes that happened within a certain distance from and a reasonable period after another crash. In a time-space diagram originated from the primary crash (Figure 2), such criteria form a rectangular window and any crash falling in that window is considered a secondary crash. Raub was among the earliest researchers who tried to identify secondary crashes in this way *(9)*. The definition of a secondary crash given by Raub was a crash that occurred within 1600 meters upstream of an earlier incident and within 15 minutes after the clearance of that incident. A case study was conducted on arterial roadways in seven continuous urban areas of Northern Chicago, Illinois. The proposed secondary crash identification method was simply a filtering by the definition based on data sources including police response to crashes, traffic violations, disabled vehicles, and other incidents on arterial roadways. The data consisted of 1796 incidents, of which 627 were crashes, 44 were fires, 534 were traffic stops, 491 were disabled vehicles, and 100 others. The study found 97 secondary crashes for 81 primary incidents. Additional analyses showed average time separation between a secondary crash and the primary incident was 36.4 min. Forty percent of the secondary crashes occurred within 20 min of the primary incidents and more than 55 percent happened within 30 min. Average spatial separation between a secondary crash and the primary incident was 0.4 miles (600 m). About 35 percent of the secondary crashes occurred within 200 meters of the primary incidents and more than 50 percent occurred within 600 meters. When large spatial separation occurred, it was either because more than one secondary crash happened or because the duration of the primary incident was larger than 60 min. Raub's method had a major influence on some following research *(4, 10, 11)*. For example, Karlaftis et al. (1999) made a minor modification to the space threshold by adjusting it to 0.8 km (later increased to 1.5 km), and identified 35 percent of the accidents on the Borman Expressway from 1992 – 1995 as secondary crashes *(4)*.

Concerns have been raised that using the same spatial and temporal thresholds for different primary incidents ignores the variation in location types, traffic conditions, etc. Zhan et al. proposed using the maximum traffic queue length and the maximum queue dissipation time as the spatial and temporal thresholds, respectively, for each primary incident *(12)*. They used hourly traffic volumes and incident duration information to estimate the maximum traffic queue length resulting from the primary incident; then, the maximum queue dissipation time was computed accordingly. Any crash occurring within the boundary of the estimated maximum queue length and dissipation time was then classified as a secondary crash to the corresponding primary incident. To apply this method, traffic counts and incident data were both used. The incident data in the study covered a 2-year period from January 2005 to January 2007 for three major freeway corridors managed by the Florida Department of Transportation in Fort Lauderdale. The total number of incidents was 95,844, with 7903 of those incidents being crashes. For each incident, the start and closure times, geographical information, environmental conditions, incident type and severity, lane and shoulder blockage information, number and type of vehicles involved, incident response status, and other information were detailed. After identifying the secondary crashes, descriptive, statistical, and logistic regression analyses were applied to identify potential factors that contributed to the secondary crashes. As a result, for general cases, the primary incident type, primary incident lane-blockage duration, and time of day were found to be significant when predicting the likelihood of a secondary crash. For the specific study, incidents occurring on northbound I-95 were also identified as a significant factor. Although maximum queue length and maximum dissipation time varied among incidents, they were still static thresholds with respect to a certain incident.

In spite of treating each primary incident differently, Zhan's approach still belongs to static thresholds in the sense that the spatial threshold does not change over time. In reality, a queue caused by an incident grows and discharges. Static thresholds ignore such reality and may result in a false secondary crash that did not occur in the queue. In addition, if the chosen temporal threshold does not cover the entire duration of queuing, a true secondary crash could be missed. As a solution to these problems, researchers proposed a family of dynamic methods.

***Dynamic queue end***

Moore et al. gave a refined definition of secondary crashes and incorporated the dynamic characteristics of queues in their identification method *(5)*. They argued that a secondary crash only occurs within the boundary of a high density queue caused by an initial incident and that chain reaction crashes should not be considered secondary crashes, because even if studied, they could not be prevented effectively by patrol service; also, a secondary crash can only happen upstream from the initial incident location in either direction. In their study, an initial incident could be any type, not only a traffic accident. Additionally, one initial incident could be responsible for more than one secondary crash.

Under this definition of a secondary crash, Moore et al. performed a detailed investigation to develop accurate estimates of secondary crash rates in the Los Angeles area. Data were collected from the First Incident Response Services Tracking (FIRST) warehouse and the Caltrans District 7 loop detector data. A total of 84,684 FIRST records were retrieved covering 393 center-line miles of freeways in Los Angeles County for the months of March, May, and July in 1999, plus the last week of December of 1998. Each FIRST record provided detailed information for an incident, including the location on the freeway and the nature of the incident. Duplicates existed among the records. The loop detector data consisted of the freeway volumes and speeds, 24 hours a day, seven days a week. Based on the data quality and the definition of secondary crashes, the authors developed a 4-filter method to identify primary/secondary accident pairs. First, incidents were excluded if they occurred outside the 2-hour-2-mile boundary centered by any other incident. The 2-hour-2-mile criterion was obtained through field observation of queues. After the first filter, 31,259 incidents remained as candidate primary/secondary pairs. Second, a potential secondary incident was excluded if it was not a crash, happened downstream of the primary incident, or was qualified as a chain reaction crash. The second filter preserved only 1,078 incidents (529 candidate pairs) for further filtering. The third filter eliminated duplicates, leaving only 389 incidents (192 pairs) for the last step. The last filter used the detector data to calculate shock-waves and determined if a secondary crash actually occurred within the boundary of a high density queue caused by the primary incident. As a combination effect of the last filter and other data quality issues, only 5 secondary accidents were confirmed, which possibly represented just a small portion of actual secondary crashes.

Based on the filter procedure, the suggested range of the number of secondary accidents per accident was 0.015 to 0.030 and the range of the number of secondary accidents per incident was 0.007 to 0.013. This study set up a step-wise procedure to identify secondary crashes.

Careful constraints after the first filter helped to overcome the limitation of using static thresholds. However, imposing queuing conditions on secondary crashes is too strict, because some secondary crashes were caused by conditions other than queuing (e.g., distraction of the clearance). In general, the proposed filter method was considerably complex and had strong dependency on the data type and quality.

With a similar definition of secondary crashes as Moore et al., Sun et al. proposed the concept of a progression curve to identify secondary crashes *(13)*. A progression curve would represent the incident queue length as a function of time, starting from the occurrence of the incident (as illustrated in Figure 2). A crash would be considered a secondary crash of a primary incident if the crash fell under the primary incident's progression curve. To develop a progression curve, the authors used intranet traffic reports that kept track of the end of the queues throughout an incident. Complex and intensive programming work was needed to convert the intranet records into a useful format for curve plotting. In a case study using the proposed method, the author collected 5,514 accidents from I-70 and I-270 in Missouri during 2003, the police reports for which were available and contained essential information like direction, continuous log, date, severity, time, and traffic condition. For the day time periods, the progression curve method found 321 secondary crashes, while the traditional static thresholds identified 313 secondary crashes. For the all-day periods, the progression curve method found 397 secondary crashes, while the traditional static thresholds identified 390 secondary crashes. Although the two methods estimated equivalent numbers of secondary crashes, the traditional method only captured about 68 percent of secondary crashes identified by the progression curve method. In other words, if the progression curve method captured all true secondary crashes, then the traditional method accepted a large number of non-secondary crashes while rejecting an equivalent number of actual secondary crashes. Although the progression curve method was attractive because of how it addressed the relationship between time and space thresholds, it was hard to implement due to the difficulty of retrieving useful intranet reports and format converting. Additionally, the paper did not explicitly generalize the progression curve to both directions of traffic, and still suffered from missing secondary crashes that were not caused by incident queues.

Orfanou et al. developed a dynamic spatiotemporal threshold method to identify secondary crashes, based on real-time traffic data and upstream detector data in freeways *(14)*. The basic idea of Orfanou's method is similar to Sun's progression curve, but instead of using intranet data and supporting media data to track the boundaries of the queue, Orfanou et al. used an ASDA model to estimate the queue fronts based on real-time traffic flow and upstream detector data. A case study was conducted on a toll way with cameras managed by the Traffic Management Center and loop detectors placed every 500 meters on open roads and 50 meters in tunnels. Historical data for 856 accidents in 2007 were used. Secondary crashes identified by the proposed method were different compared to those identified using static thresholds.

Using the dynamic queue end as the boundary of secondary crashes is more realistic than using static thresholds, but the queue end is not the only boundary of a high density queue. After the primary incident is cleared, the queue starts to discharge. Two shockwaves exist at the same time, the queuing shockwave and the discharging shockwave. The queuing shockwave continues pushing the queue end further upstream while the discharging shockwave pushes the queue head upstream at a greater speed. The length of the queue decreases as the queue head catches up with the queue end and the queue diminishes when the two ends meet. According to the definition, a secondary crash that happened in the queue should be bounded by both the queue head and the queue end. The area in the time-space diagram bounded by these two shockwaves is called the impact area in this study.

***Impact area***

A commonly seen approach to locate the impact area for secondary crash identification is by observing speed reductions. Chou et al. is among the earliest to use the impact area for secondary crash identification *(15)*. The basic idea of their method was to find an area on a time-space diagram within which the average speed of upstream traffic of the incident was reduced by a predetermined percentage. Such an area was defined as the impact area. Crashes that fell within the impact area were considered secondary crashes. The representative boundaries of different impact areas under different traffic conditions and incident properties (e.g., lane blockage, clearance time, etc) were simulated through CORSIM. Regression models were then developed to predict the four corner points of an impact area. For each given historical incident, the impact area could be estimated using traffic data and incident properties via the corner point models. A case study was performed using 693 primary incidents that occurred along a 10 mile segment of I-287 (without ramps) in the state of New York. Comparison with traditional static time space thresholds shows a significantly reduced misclassification rate of 58 percent and greater. In spite of the reduced computational effort required by the original method of Haghani et al. *(16)*, considerable CORSIM modeling, calibration, and simulations are still required for any new study site. Additionally, visual identification is needed in collecting the coordinates of simulated corner points. This could introduce human errors into the regression data, which adds to regression errors. As a more direct application of the speed reduction based approach, some researchers simply used the actual speed data to capture the impact area *(17, 18)*.

Finding the impact area using speed reduction as the indicator is conceptually ideal, but it has several drawbacks. First of all, highly detailed speed (or travel time) data are required for relatively small spatial and temporal intervals so the time-space diagram can be approximated by a time-space grid. However, densely deployed detectors are not available on most freeway systems. Even with a recent attempt to use mobile-device-sourced traffic data, accuracy of the derived speeds are not satisfactory at a desired temporal resolution *(19)*. Second, even with high resolution data, the resulting impact area could be sensitive to the selected level of speed reduction. In other words, an impact area with 50% or more speed reduction could be very different from an impact area with 60% or more speed reduction. The choice of speed reduction percentage is not well defined. Third, as mentioned above, the time-space diagram can only be approximated by a time-space grid, which could not reflect the continuous and smooth change of queues. Last but not least, the computation effort is heavy and not suitable for applications on large scale networks.

### 2.1.2 Influential Factors on Secondary Crashes

In the literature, researchers have found a number of factors that were related to the likelihood of secondary crash occurrence. A summary is given below in Figure 3, followed by a more detailed listing of all the related studies.

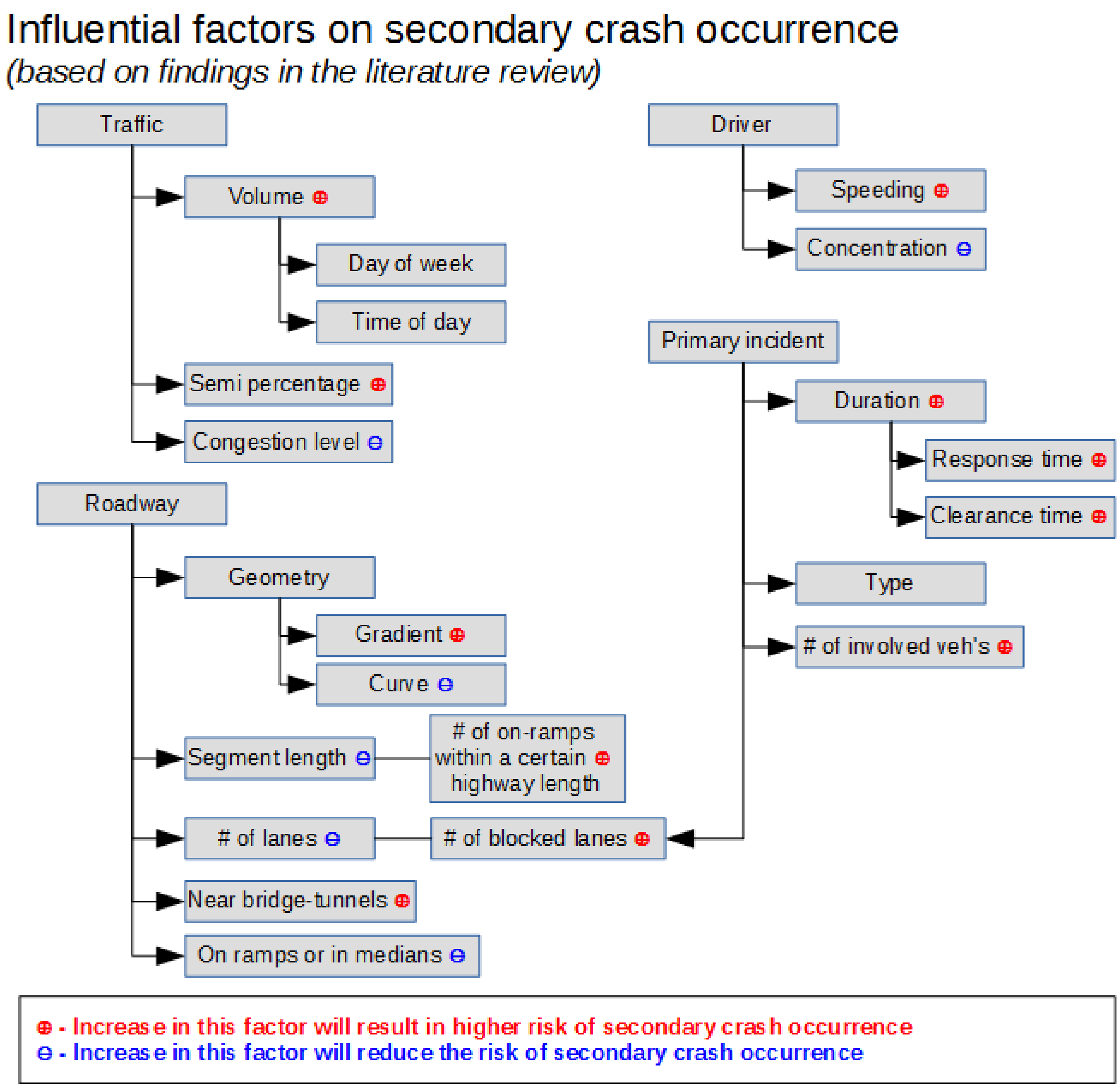



**FIGURE 3 Summary of influential factors on secondary crash occurrence in the literature.**

Owens was among the earliest researchers who observed secondary crash causes *(20)*. Owens conducted an on-site spot study on traffic incidents occurring on a 21 kilometer segment of the M1 motorway in Hertfordshire, United Kingdom. Police patrol cars observed a total of

180 hours of traffic during peak periods (flow exceeding 3500 veh/h) between July and September. They observed traffic incidents, incident effects on traffic behaviors, and police clearance procedures. Observations were recorded through written logs and sometimes with a tape recorder for rapid incidents. Based on the observations, the author identified 13 crashes as secondary crashes of which the primary causes were determinable: 3 secondary crashes happened because the drivers were distracted by police clearing the primary incidents; the other 10 secondary crashes occurred within or at the end of queues formed from the primary incidents. The author concluded that the causes to the observed secondary crashes were a combined effect of high traffic flows, capacity constraints, poor design features, competitive driving conditions, and an easily distracted driver population.

Karlaftis et al. (1999) examined the Intelligent Transportation System (ITS) impacts on safety and traffic management, with particular focus on investigating the contributing factors to secondary crashes *(4)*. Incident records on a 16 mile stretch of the Borman Expressway during 1992 - 1995 were provided by the Indiana Department of Transportation. Five variables were chosen in modeling the likelihood for a primary crash to cause a secondary crash: clearance time, vehicle type, lateral lane position, season, and weekday or weekend. Two logistic models were selected, one assuming all five variables to be independent, the other treating clearance time to be related to season. Regression results showed that the likelihood of a primary crash followed by a secondary crash increased as the clearance time increased; the likelihood was also higher when cars and semi-trucks were involved; weekdays, implying higher traffic volumes, also resulted in higher chances of causing a secondary crash; primary crashes happening on ramps or medians had a lower probability of being followed by a secondary crash. The first model, counter-intuitively, suggested that secondary crashes had lower likelihood in winter. Further based on the odd ratio of each variable, the authors estimated the monetary benefit achieved by the Hoosier Helper patrol via reducing secondary crashes, which proved to be significant. A practical advantage of this paper is that the five chosen independent variables are commonly available from most crash records and sometimes imply other typical aggregated factors (e.g., weekday or weekend is a surrogate measurement of traffic volume), facilitating the use of the likelihood models.

Hirunyanitiwattana et al. investigated the secondary crash characteristics based on crash records from the Highway Safety Information System (HSIS) *(10)*. According to crash records covering the California highway system, 170,520 primary crashes and 7,338 secondary crashes occurred in 1999 and 183,641 primary crashes and 8,104 secondary crashes occurred in 2000 (based on a 60 minute and 2 mile static window). Comparison analyses showed that urban districts had higher probabilities of secondary crashes than rural districts. The distributions of secondary crashes differed from that of primary crashes with respect to time of day, crash severity, collision type, primary collision factor, and road classification. Speeding was found to be the highest collision factor for both primary and secondary crashes.

Khattak et al. tried to explain the deeper relationship between incident durations and secondary incidents *(3)*. Data for 38,086 incidents were collected from the Hampton Roads Smart Traffic Center in Virginia Beach, Virginia for the year of 2006. The data covered the freeways patrolled by the Safety Service Patrol of the Hampton Roads area. Prediction models were developed for incident duration and secondary incident occurrence. The hypothesis about the interdependence between the primary incident duration and secondary incident occurrence was tested and proved to be statistically significant.

Zhan et al. applied both descriptive statistical analyses and logistic regression analyses to crashes to identify potential factors that contributed to the secondary crashes *(12)*. As a result, for general cases, the primary incident type, primary incident lane-blockage duration, and time of day were found to be significant when predicting the likelihood of secondary crashes. For the specific study case, an incident occurring on northbound I-95 was also a significant factor.

Zhang, et al. lead a study on multiple secondary incidents *(21)*. A detailed freeway incident data set was obtained from the Traffic Operation Center in Hampton Roads, Virginia for the year of 2005. The results indicated that longer duration crashes, shorter segments, and heavy traffic were associated with higher propensity for secondary incidents. Further, multiple-vehicle involvement and lane blockage were associated with multiple secondary incidents.

Vlahogianni et al. investigated the effects of traffic conditions and incident characteristics on the upstream influence area of a crash *(22)*. A total of 1,746 crash records for 2007 and 2008 were collected on a 65.2 km stretch of an urban motorway, Attica Tollway, in Athens. A Bayesian network for the probabilistic estimation of different influence areas for secondary crash occurrence was developed with respect to various incident and traffic characteristics. Results indicated that the upstream influence area of a crash was significantly determined by the traffic conditions and incident response and clearance time.

Khattak et al. examined the spatial relationship between secondary and non-secondary incidents as well as factors of roadway segments that resulted in high secondary incident frequency *(23)*. Road inventory data for 2006 were obtained from Hampton roads Planning District Commission, and the traffic flow data were provided by the Virginia Department of Transportation. A total of 41,539 incidents in 2006 were collected for five major interstate freeways in Virginia. Through regression modeling, the authors found that the spatial distribution of secondary incidents was different from that of non-secondary incidents, while places with high non-secondary incident frequencies did not necessarily have high secondary incident frequencies. Also, long roadways, high traffic volumes, and more on-ramps on freeway segments were found to be associated with higher secondary incident frequency, while segments with curves, more lanes, and higher congestion levels were associated with lower secondary incident frequency. When roadways were close to bridge-tunnels, both non-secondary and secondary incident frequencies were high.

Orfanou et al. found that some determinants of the primary crashes, such as the initial traffic condition, the number of blocked traffic lanes, and the number of vehicles involved in the crash, significantly affect the spatiotemporal evolution of the primary crash, and hence the likelihood of a secondary crash occurring *(14)*. Vlahogianni et al. *(24)* expanded this study to include real-time weather information, but did not find a significant impact from weather.

## 2.2 Traffic Incident Management

Traffic incident management (TIM) is part of emergency transportation operations (ETO). The Federal Highway Administration (FHWA) defines TIM as "a planned and coordinated multi-disciplinary process to detect, respond to, and clear traffic incidents so that traffic flow may be restored as safely and quickly as possible" *(25)*. When an incident occurs, it can slow down the traffic by blocking certain lanes. Such unexpected slowing or stopping leads to sudden changes in speeds, which is a well-known contributor to freeway crashes *(26)*. A crash can also happen due to drivers being distracted by the incident scene *(20)*. Most secondary crashes fall in one or both of the above two cases. Not only will the incident scene cause safety concerns for moving traffic, but the moving traffic can also impose safety concerns for the incident responders, leading to struck-by crashes. Effective TIM shortens the influential period following a traffic incident and reduces safety risks and the congestion duration (if any) along the highway. Researchers and practitioners have conducted and published numerous programs and studies on how to evaluate and improve TIM.

### 2.2.1 Who and What Are Involved in Traffic Incident Management

TIM involves incident detection, incident response (or treatment), and incident clearance, and successful TIM strategies highly depend on effective cooperation between multiple agencies. Traditional participants in a TIM system include, but are not limited to, law enforcement agencies, fire and rescue agencies, emergency medical services, transportation agencies, public safety communications, emergency management agencies, towing and recovery agencies, hazardous material contractors, and traffic information media *(25)*. The typical agencies and their responsibilities during TIM are summarized in Table 1. Zhou, et al. gave a summary of some relevant TIM agencies in 15 states *(27)*.

## TABLE 1 TIM RELATED AGENCIES AND RESPONSIBILITIES

| **Agency Type** | **Typical agencies** | **TIM Stages** | | |
|---|---|---|---|---|
| | | Incident Detection | Incident Response | Incident Clearance |
| Law Enforcement | *State Police, Highway Patrol, County Police/Sheriffs, etc* | Detect incidents directly at their occurrence while patrolling | Secure, investigate, provide emergency aid, etc | Supervise and direct traffic |
| Fire and Rescue | *Fire department* | | Protect scene, suppress fire, provide emergency medical care, rescue victims, etc | Assist |
| Emergency Medical Services | *Fire and rescue companies* | | Provide advanced emergency medical care, coordinate evacuation, etc | |
| Transportation | *Traffic Operation Center (TOC)* | Assist in detection and verification | Initiate traffic management strategies, provide traffic control, provide motorist information, etc. | Coordinate clearance, provide repair needs, repair infrastructure, etc |
| Public Safety Communications | *911 call takers and dispatchers* | Receive incident notifications | | |
| Emergency Management | *Special duty agencies in state and local governments* | | Response to large-scale emergencies | |
| Towing and Recovery | *Towing and recovery service providers* | | | Recover and remove vehicles from incident scene, protect victims' property, remove debris from the roadway, etc |
| Hazardous Material Contractors | *Hazardous & waste contractors* | | | clean up and dispose of toxic or hazardous materials |
| Traffic Information Media | *Private traffic information vendors* | | Broadcast traffic information at incident scene through different ways, such as radio, internet, etc | Broadcast recovering traffic information |

### 2.2.2 How to Evaluate the Performance of Traffic Incident Management

The performance of TIM is a combined result of multiple factors. A typical example is incident clearance time. According to a number of research studies, the faster an incident is cleared, the lower the risk of safety issues (e.g., secondary crashes) becomes *(3, 4, 22, 28)*.

In 2002, under contract with FHWA, the American Transportation Research Institute (ATRI) developed a Traffic Incident Management Self-Assessment (TIMSA) questionnaire to measure TIM performance and identify potential program gaps at the national level *(20)*. Since then, nationwide TIMSA has been conducted annually, with slight modifications between 2008 and 2011. The aspects considered important by FWHA in evaluating TIM performance can be perceived through 2009 TIMSA question categories, as demonstrated in Table 2 *(29)*.

**TABLE 2 2009 TIMSA QUESTIONNAIRE OVERVIEW**

| Categories | Subcategories | Number of Questions |
|---|---|---|
| Strategic (formerly Program and Institutional Issues) | Formal TIM Programs | 2 |
| | Multi-Agency TIM Teams | 4 |
| | Performance Measurements | 5 |
| Tactical (formerly Operational Issues) | Policies and Procedures for Incident Response and Clearance | 7 |
| | Responder and Motorist Safety | 5 |
| Support (formerly Communications and Technology Issues) | Data Collection/Integration/Sharing | 5 |
| | Traveler Information | 2 |

### 2.2.3 How to Improve Traffic Incident Management System

Countermeasures to improve TIM systems correspond closely to the factors used in the performance evaluation step. Zhou, et al. used communication, coordination, and cooperation (3 C's) to summarize the programs they proposed for development of TIM *(27)*. FHWA provides a Second Strategic Highway Research Program (SHRP 2) to promote shared understanding of the requirement for on-scene management and quick clearance. Carson summarized critical challenges and proposed some strategies in incident detection and verification, traveler information retrieving, incident response, scene management and traffic control, and quick clearance and recovery *(30)*.

In 2003, the National Cooperative Highway Research Program (NCHRP) published the NCHRP Synthesis 318 Report to provide to the incident management agencies the guidelines on quick and safe clearance of roadway traffic incidents *(7)*. Freeway service patrol was suggested to function effectively in detection, verification, response, and removal of minor incidents.

The Traffic Incident Management Handbook *(6)* specifically listed the following objectives as program-level performance improvements: 1) reduce "road clearance" time, 2) reduce "incident clearance" time, and 3) reduce the number of secondary crashes.

### 2.2.4 Related Works

Karlaftis et al. *(4)* built two logistic models to predict the likelihood for a primary crash to be followed by a secondary crash based on incident data compiled from Indiana's Hoosier Helper Freeway Service Patrol program daily activity log. According to the paper, the Hoosier Helper program was initiated in September 1991, supported by the Indiana Department of Transportation. At the time of the paper, the program maintained a fleet of six vehicles, patrolling over a 16 mile stretch of Interstate 80-94 near Gary (known as Borman Expressway) and an 8 mile stretch of Interstate 65. The patrol services included providing support at crash sites, supplying fuel, changing flat tires, and calling tow truck operators. A daily activity log was written to document all responses. Based on the likelihood models and corresponding odd ratios, the existence of the Hoosier Helper program helped to reduce secondary crashes in the amount of approximately $568,080, exceeding the 1995 cost of the program itself. Further cost-effective analysis was conducted by Latoski et al. *(31)* to evaluate the Hoosier Helper program's benefit more comprehensively. All calculations were done for each of the two scenarios: daytime patrol and 24-hour patrol. Agency cost was assumed to consist of annual investment cost, employee salaries and benefits, overhead costs, and maintenance costs. Benefits were assumed to consist of non-recurrent congestion delay savings, secondary crash reduction, and vehicle operation cost savings. The results showed a 4.71:1 benefit-cost ratio for daytime operations and a 13.28:1 ratio for 24-hour operations.

Guin et al. *(32)* analyzed the benefits for an incident management program integrated with intelligent transportation systems operations. A methodology was developed to compute benefits from motorist assistance services, the reduction in delay, fuel consumption, secondary crashes, and the improvement in air quality. The Georgia NaviGAtor system was used in the case study, as it was a highly integrated transportation management system that used a variety of technologies and processes to monitor the operation of the freeway and arterial system, respond to a variety of incidents, and disseminated traveler information. Non-recurrent traffic congestion caused by incidents was expected to be reduced by NaviGAtor, improving overall highway efficiency. The first strategy implemented by the Georgia Department of Transportation in order to fulfill the goals of the Traffic Management Center was a freeway management and advanced traveler information system for collecting traffic information, managing incidents, and disseminating information to drivers and news media outlets. The proposed methodology computed an annual savings of 7.2 million vehicle hours of incident-related delay and $187 million savings for a 1 year period during 2003 and 2004. The annual benefits-cost ratio was calculated to be 4.4:1 for the NaviGAtor system.

An attempt to keep track of secondary crashes was done in Kentucky by adding a "Secondary Collision" field to the Uniform Police Traffic Collision Report *(8)*. Although the accuracy was not found to be satisfying, further education and training for the police officers were suggested to enhance the quality.

Ng et al. proposed a Semi-Markov process based model to predict the time to primary incident and the time to secondary incident *(33)*. Their idea was to incorporate the stochastic feature and the dynamic feature of traffic to model the state change of highways (between normal state, primary incident state, and secondary incident state), thus allowing TIM managers to make better decisions regarding resource dispatching. A numerical experiment of this model suggested that the time to secondary incident could be increased if the primary incident was cleared and the highway was restored to a normal state. However, this model was limited to the analysis of a highway segment. It was recommended as future research to expand the model onto a highway network.

## 2.3 Freeway Service Patrols

A freeway service patrol (FSP) is an umbrella term for a group of freeway service programs implemented by state departments of transportation and highway patrols, dated as early as 1960 *(34)*. The general mission of FSP is to alleviate traffic congestion and maintain highway safety through patrolling designated highway sections, providing assistance to motorists in trouble, and performing incident management. The formation and operation of an FSP program can vary from state to state. Differences may exist in numbers and types of patrol vehicles, patrolling scopes, patrolling hours, etc.

Cost-benefit analysis is an important reference for future implementation of a new FSP program or expansion of existing FSP programs. A number of recent studies in the literature continue to show the benefits gained by FSP programs as low-cost measures under the intelligent transportation system (ITS) architecture. For examples, the Freeway Incident Response Safety Team (FIRST) of the Minnesota Department of Transportation was reported to generate $16.62 million in monetary benefits in 2003 *(35)*. Compared to an estimate of $2.73 million in 2000, the large increase was attributed to both the expansion of the program and the inclusion of emission measures and secondary crash reduction in benefit calculation (in addition to delay and fuel consumption). As of 2005, the Road Ranger service patrol in Florida showed an overall benefit-to-cost ratio of over 25:1, ranging from 2.3:1 to 41.5:1 for the six studied sites *(36)*. An updated study reported the benefits (delay and fuel saving only) of the Road Ranger program to be about $135 million in 2010, with the cost being $19.9 million and benefit-to-cost ratio around 6.7 *(37, 38)*. In Virginia, average incident duration was estimated to be reduced by approximately 17% with the support of the Northern Virginia Safety Service Patrols, contributing to an overall benefit-to-cost ratio of 6.1:1 *(39)*.

In spite of positive net benefits, the underlying cost of FSP programs cannot be ignored. Limited patrol resources (i.e., patrol vehicles, operators, equipment, etc.) are a major concern for most FSP programs. An efficient allocation of patrol vehicles that focus on hazard and congestion prone freeway sections may help to increase the utilization of FSP resources, as well as maximize the highway system performance. An example is Yin's study on optimizing the allocation of tow trucks on freeway networks to minimize the worst-case system travel time during incidents *(40)*. In the current study, freeway locations with a statistically higher risk of secondary crash occurrence should be identified and included as hotspots for FSP managers.

## 2.4 Summary

A comprehensive literature review was performed on previous secondary crash studies, traffic incident management, and freeway service patrols.

Existing definitions and identification methods of secondary crashes were found and compared, as well as previous findings on the influential factors on secondary crash occurrence. As far as the literature is concerned, few state-level secondary crash studies can be found, leaving a research gap in establishing an efficient and reliable procedure to identify secondary crashes on a large scale highway network. Such a procedure should consider incorporating the dynamic spatial and temporal relationship between a primary crash and its secondary crash for more accurate identification; on the other hand, the procedure should strive to minimize its complexity so when it is applied to a large scale network, reasonable efficiency can be achieved. In terms of verification, most previous studies did not manually examine crash reports to validate their status as secondary crashes, weakening the validity of the analyses (e.g., the influential factors). Some studies validated secondary crashes, but the resulting sample sizes were too small due to the study scope, preventing further investigation. Thus, for the current project, the proposed procedure should also contain manual review to validate secondary crashes; and since a statewide network over a long term period will be studied, the resulting final sample size is expected to be relatively large.

Review on TIM provides a high level understanding of involved partners, TIM performance evaluation methods, TIM improvement strategies, and some supporting research efforts. Review on FSP shows the purpose and benefit associated with FSP deployment. More importantly, practical needs for the current project were revealed through the above understandings. In order to provide the state TIM agencies with useful information for reducing secondary crashes, historical data should be used. Lacking a flag for secondary crashes in most crash databases is a major obstacle. Development of a procedure that can easily and quickly identify secondary crashes from the historical data would bridge the gap between the practitioners and the latent information in the crash archives. In addition, with secondary crashes identified from the statewide highway network, incident hotspots can be located to guide the deployment or expansion of freeway service patrols. The spatial-temporal relationship between primary incidents and secondary crashes can help freeway service patrols develop better searching ranges for secondary crashes upon the detection of a primary incident.

# CHAPTER 3 SECONDARY CRASH IDENTIFICATION METHOD

In this study, ***a secondary crash is defined as a crash that occurred due to a non-recurrent highway situation caused by a primary incident, including congestion and distractive or hazardous incident scene; a secondary crash can happen upstream of, downstream of, or at the initial primary incident location in both traffic directions***. Although the above definition encompasses all primary incident types, the current study would only have crashes for primary incidents due to data availability. Additionally, deer-related and weather-related hazards would not be considered in the study scope, but left as future work. As noted by Moore et al., chain reaction crashes have little significance on secondary crash investigations, since they followed the initial incident so rapidly and cannot be prevented through TIM *(5)*. In addition, a chain reaction crash could have been documented together with the initial incident as one single crash, making it hard to separate from the historical data. Thus, chain reaction crashes are not considered as secondary crashes in this study.

***A secondary crash can only be confirmed when another crash is mentioned in its crash report as a causal event***. Such criterion may result in false negatives where a true secondary crash is missed due to the lack of mentioning another crash, but it ensures that the identified secondary crashes satisfy the definition given above. Another important consideration of using crash reports as the end reference is that they are informative enough to distinguish a chain reaction crash from a secondary crash. From the description of the crash report, one can identify a chain reaction crash if it was an immediate collision with any vehicle involved in the previous crash. Chain reaction crashes normally have zero time difference and zero distance from the previous incident; but using these spatial and temporal criteria to eliminate chain reaction crashes is not reliable, because due to data precision, two crashes (e.g., a true secondary crash and its primary crash) could be documented with identical accident time and location while they were actually several vehicles apart and several minutes away.

Directly reviewing all crash reports to confirm secondary crashes is so time- and labor-consuming that some reasonable, fast, and automatic procedures are needed to reduce the amount of candidate crash reports. Based on techniques explained in the literature and the availability of data, the research team developed a multi-phase method that reduces the massive statewide crash data down to a small subset for secondary crash confirmation. The method is designed for crash data that can be located on the highway network using a linear referencing system. In general, the method consists of three major stages (Figure 4). The first stage is a filter for excluding crashes outside the scope of interest. The second stage explores the spatial-temporal closeness among the rest of the crashes to identify crash pairs that do not exceed certain thresholds for time difference and network distance. In the third stage, filters are applied to each pair generated from the second stage. The resulting pairs are considered candidates of primary-secondary crashes and are reviewed by crash report reviewers. Although the number and the order of stages are fixed, each stage is designed to be a flexible module, i.e., whose filtering or processing criterion can be customized for better accuracy or specific research needs. Figure 4 shows the criterion used in each stage of the current study. The next subsection describes the available raw data, followed by three successive subsections elaborating on each of the three stages. Some additional description about the manual review process is given in the last subsection.

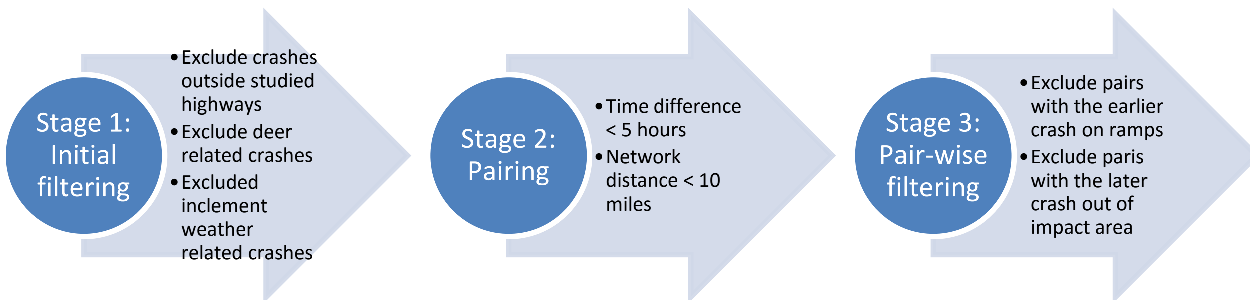



**FIGURE 4 An automatic method to find secondary crash candidates.**

## 3.1 Raw Data Description

A general knowledge about the primary data source is helpful for understanding the design and implementation of the three-stage method for secondary crash candidate extraction. The WisTransPortal Data Hub of the Traffic Operations and Safety (TOPS) Laboratory at the University of Wisconsin – Madison houses a variety of statewide transportation data prepared and provided by the Wisconsin Department of Transportation (WisDOT) *(41, 42)*. Among these data, the State Trunk Network (STN) data and the crash data are the two primary sources for the current project. The STN includes the state trunk highways (STH), the US highways (US), the interstates highways (IH), the designated freeways, and the expressways in Wisconsin as of 2012 *(43)*. The crash data cover all reported crashes in Wisconsin since 1994 and are updated monthly. WisDOT provides both maps (i.e., Esri™ shapefiles) and database records of the STN and crashes. Maps of the STN are maintained in two formats: chains and links (Figure 5). STN chains are polyline features in ArcGIS, each aligning on the centerline of the represented directional highway segment. STN chains are suitable for visual demonstration and spatial selection. STN links are straight line features in ArcGIS, each also representing a directional segment of highway, but only consistent with the highway geometry in terms of the start point and end point positions. One way to locate a crash on the STN is to use the crash map and the STN chains in ArcGIS. Each crash is represented as a point feature with latitude-longitude coordinates. If a crash happened on the STN, it should be mapped almost on an STN chain in ArcGIS, at worst several meters away. By using a spatial selection with a small buffer distance, all these STN crashes can be correctly identified. As will be mentioned later, this method is used in the initial filtering of input crashes. Another and even more important way of crash locating for this project is to use linear referencing based on STN links. WisDOT embeds a linear referencing system in the crash records to allow a crash to be located on the STN links without a planar search. Details of the linear referencing system are given below.

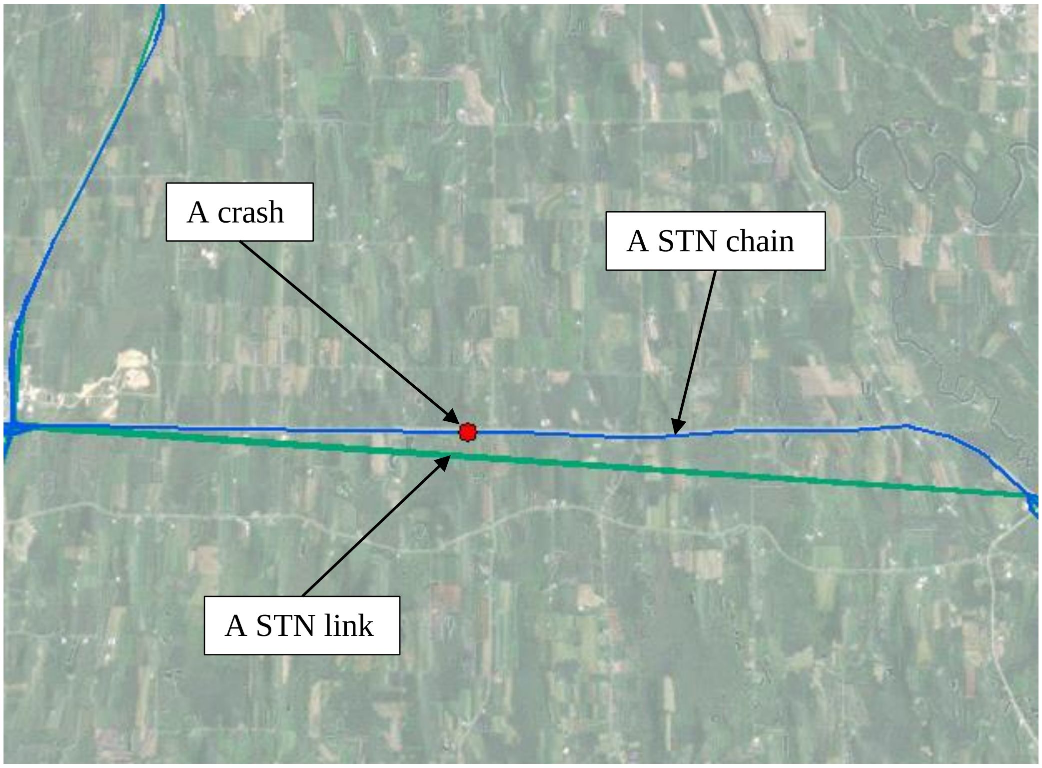
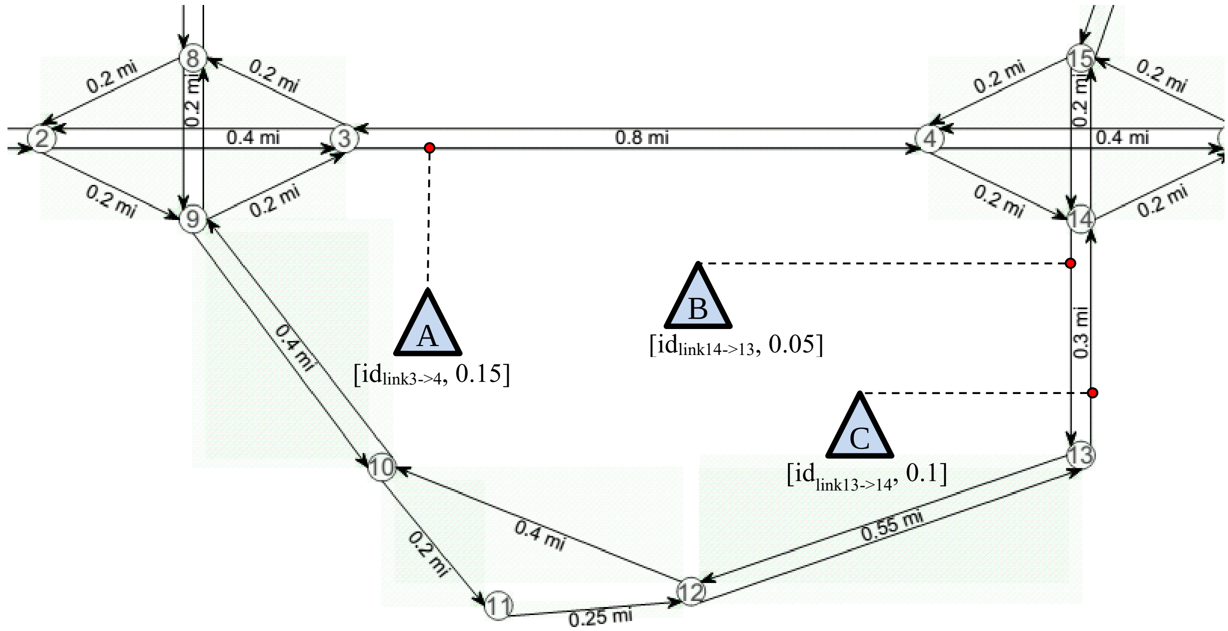



**FIGURE 5 Illustration of STN chains, links, and crashes in ArcGis.**

### 3.1.1 STN Links and Linear Referencing

A traditional way of modeling highway networks is a figure that consists of nodes and directional links. The STN links (or "links" hereafter) are used in this manner. Nodes in the STN are called reference sites (RS). Each link starts from one RS ($RS^{from}$) and ends at another RS ($RS^{to}$). A link normally represents a highway segment, either mainline or ramp, in one traffic direction. Figure 6 illustrates how a small portion of a highway network is represented by links. The links are displayed as solid arrows with their lengths. RS's are labeled in circles. An arbitrary location on the STN, such as a crash scene, is determined by a linear coordinate [*link ID*, *link offset*], namely linear referencing. The *link ID* identifies the link on which the crash occurred and the *link offset* is the distance from the link's $RS^{from}$ to the crash. As of the year 2012, the total number of in-operation links is 33,015 and the total length is 24,903 miles. Note that in Figure 6, the marked link lengths are actual highway segment lengths instead of the straight line lengths. Similar distinction should be made between the link length and the feature length in ArcGIS. The link length is always larger or equal to the straight line length.




***Example distances:***

*d(A,B)*= *(*0.8 – 0.15) + 0.2 + (0.3 – 0.05) = 1.1 miles

*d(B,C)* = 0.3 – 0.05 – 0.1 = 0.15 miles

**FIGURE 6 Illustration of the linear referencing system.**

### 3.1.2 Crash Records

WisTransPortal stores each police-reported crash in Wisconsin as a record in the database. Each crash record contains detailed information about the crash, such as a unique identification number, date, time, link ID and link offset of the crash location (for linear referencing). Also, each crash record is associated with a document ID that links the crash to its police report form, called MV4000 in Wisconsin. The MV4000 form provides additional information such as the police narrative with a crash diagram.

### 3.1.3 Other Data

In addition to the STN links, WisTransPortal also stores other highway information. For example, all routes in the STN are stored in a table, each record representing the entire stretch of a highway route and its geographical direction (e.g., US 12 East Bound); virtual mile markers are stored as reference points. Traffic data are also available. WisDOT manages the TRaffic DAta System (TRADAS) and the Advanced Transportation Management System (ATMS), with traffic detectors deployed on the STN. WisTransPortal contains information of these TRADAS and ATMS detectors, as well as their traffic counts.

## 3.2 Stage 1: Initial Filtering

In Wisconsin, the number of documented crashes by Wisconsin Department of Transportation exceeded 120,000 for the year 2012. These crashes cover all functional classifications of highways in Wisconsin, including those in parking lots and private driveways. Additionally, the collection covers all seasons with diverse weather conditions. Animal related crashes are also part of the documented crashes. The objective of this research is to study secondary crashes on access controlled highways under normal driving environments (e.g., non-inclement weather), not local roadways or in extreme driving conditions. Thus, considerable efforts can be saved by first excluding crashes that are not within the research scope. Then, the three initial filters used in the current study can be applied.

### 3.2.1 On selected STN highways

A number of access controlled highways were selected from the Wisconsin STN network as the spatial study scope (Figure 1). These access controlled highways include US highways, interstate highways, and state trunk highways. The total length of the selected highways is approximately 1,500 miles per direction of travel. Limited access and physical separation between traffic directions are two major requirements for a highway to be studied in this project. All selected highways were manually extracted based on visual verification against Google Map™, facilitated by the satellite view. The selected highways were then extracted into both a set of STN links and a set of STN chains. In order to extract crashes that happened on the selected STN highways, as mentioned previously, the STN chains were used.

### 3.2.2 Non-deer related

Deer are a common roadway hazard in Wisconsin. For example, among a total of 47,593 crashes that occurred on the Wisconsin STN in 2010, 9,615 (20%) were deer crashes. To focus the analysis on more general conditions, crashes involving deer were excluded from the current study. The DEERFLAG attribute of crash records was used to identify deer related crashes. If a crash contains “Y” for the DEERFLAG, the crash was excluded.

### 3.2.3 Non-inclement weather related

A crash was excluded if it occurred under inclement weather. In Wisconsin, a large portion of crashes were related to inclement weather during the winter. For example, in January 2010, 1,520 of 3,592 crashes (about 42%) on Wisconsin state trunk highways occurred during or after snow or rain. Some circumstances such as successive run-off-road crashes in snow storms and back-to-back rear-end crashes due to slippery or icy roads were recognized as secondary crashes. However, weather is out of the control of TIM agencies. Since the current research is focused on secondary crashes that are more likely to be prevented by effective TIM, inclement weather related crashes were not included in this study, but the authors intend to study them separately in the future. Four attributes of a crash record were used to exclude a crash involving inclement weather or unsafe road conditions. They are WTHRCOND (weather condition), ROADCOND (surface condition of the road), HWYPC1, and HWYPC2. The last two attributes both refer to the possible contributing circumstance for the highway on which the accident took place. A crash will be excluded if any of the four attributes contains "SNOW", "ICE", "WET", "SLEET", "RAIN", or "SIW".

## 3.3 Stage 2: Time-Space Pairing

Spatial-temporal closeness is a very important heuristic to further narrow down the search space after the initial filter. A standalone crash that was far from any other crash in terms of time and space could hardly be the cause or the result of another crash. In other words, only crashes that were temporally and spatially close to another crash could potentially be a primary or a secondary crash. Given the STN linear referencing system and the crash records, the target of the crash pairing algorithm is to identify all crash pairs $(c_i, c_j)$ that satisfy formulae 1 and 2. In formula 2, $d(c_i,c_j)$ is the distance between crashes measured along the roadway. Highway splits, merges, and intersections are accommodated, which was not addressed by previous studies focusing on individual freeways.

$$0 \le t(c_j) - t(c_i) \le T \tag{1}$$

$$d(c_i, c_j) \le D \tag{2}$$

where,

$c_i$ = Crash *i*, the former crash;

$c_j$ = Crash *j*, the latter crash;

$t(c)$ = The time of crash *c* since an early time origin, min;

$d(c_i, c_j)$ = The network distance between crash $c_i$ and $c_j$, mile;

$T$ = The time window (threshold), min;

$D$ = The space window (threshold), mile.

Given the significant size of the STN links and the crashes, simple algorithms are either slow or infeasible. One naïve algorithm is to run Dijkstra's method repeatedly for every crash. Dijkstra's method is an iterative approach that finds the shortest path from an origin to every node in a network. Djikstra's method can be briefly summarized as follows: All nodes are considered to be infinitely distant from the origin and "unvisited" initially. The method begins from the origin and computes the distances to its neighbors (i.e., nodes with direct connection) and marks the origin as "visited". In every successive iteration step, the method chooses the closest "unvisited" node to the origin, updates the distances from the origin to that node's "unvisited" neighbors if the paths become shorter through that node, and marks that node as "visited". The iteration continues until every node is "visited". At the end, the distances from the origin to each node are the shortest distances *(44)*. The complexity of Dijkstra's method is $O(N^2)$ with respect to *N* crashes. By repeatedly using Dijkstra's method for *N* crashes, the complexity of the naïve algorithm becomes $O(N^3)$, which is not efficient. Another alternative is to use dynamic programming to populate a shortest path matrix between every pair of crashes. This alternative is infeasible because it not only requires an amount of computation time equivalent to the first algorithm, but also requires an unmanageable amount of computer memory space (e.g., 100,000^2 * 8 bytes ≈ 75 GB) to store the matrix.

The proposed pairing algorithm first analyzes the relationships between links and uses these relationships to derive crash-to-crash distances. For each link, $lk_i$, that contains one or more crashes, the algorithm performs a variant of Dijkstra's traversal and generates the relationships between $lk_i$ and each of the other links. The distances between crashes are then calculated based on these relationships. Compared to the first algorithm mentioned above, the number of traversals is bounded by the total number of links, no matter how many crashes are analyzed. The pairing algorithm also utilizes the *D* mile space window to constrain the Dijkstra's traversal to a relevant portion of the STN. In the following subsections, the concept of a local linear coordinate system is introduced, based on which the relationship between any two links can be comprehensively defined. Additionally, the equation to derive the crash-to-crash distance from the link-to-link relationship is also given, along with the concept of a candidate link that is used to constrain the Dijkstra's traversals, the pseudo code of the algorithm with special case explanation, and finally, the validation of this algorithm.

### 3.3.1 Local Linear Coordinate System

A local linear coordinate system (LLCS) is defined for each link, namely a ***base link***, to describe the spatial relationship between any RS and the *base link*. Let $RS_{base}^{from}$ and $RS_{base}^{to}$ denote the from-reference-site and the to-reference-site of the *base link*. Under the *LLCS*, each *RS* in the network has a two-fold coordinate with the following definitions:

- **Forward (positive) coordinate ($\boldsymbol{x_{RS}^{+}}$)** = the length of the *base link* + $d(RS_{base}^{to}, RS)|\, RS_{base}^{from}$. $d(RS_{base}^{to}, RS)|\, RS_{base}^{from}$ is the shortest network distance between $RS_{base}^{to}$ and *RS* in a sub-network without $RS_{base}^{from}$ (and links connected it). If $d(RS_{base}^{to}, RS)|\, RS_{base}^{from}$ does not exist, $\boldsymbol{x_{RS}^{+}}$ = +∞. Specially, $\boldsymbol{x}^{+}_{RS_{base}^{from}}$ is defined as 0.
- **Backward (negative) coordinate ($\boldsymbol{x_{RS}^{-}}$)** = $d(RS_{base}^{from}, RS)|\, RS_{base}^{to}$. $d(RS_{base}^{from}, RS)|\, RS_{base}^{to}$ is the shortest network distance between $RS_{base}^{from}$ and *RS* in a sub-network without $RS_{base}^{to}$ (and links connected to it). If $d(RS_{base}^{from}, RS)|\, RS_{base}^{to}$ does not exist, $\boldsymbol{x_{RS}^{-}}$ = +∞. For example, $\boldsymbol{x}^{-}_{RS_{base}^{to}}$ = +∞.

As an example, in Figure 6, consider $RS_{12}$ under the LLCS of $link_{3\to4}$ (as the base link). $x_{RS_{12}}^{+}$= 0.8 ($link_{3\to4}$) + 0.2 ($link_{4\to14}$) + 0.3 ($link_{14\to13}$) + 0.55 ($link_{13\to12}$) = 1.85 miles. $x_{RS_{12}}^{-}$= 0.2 ($link_{9\to3}$) + 0.4 ($link_{10\to9}$) + 0.4 ($link_{12\to10}$) = 1.0 mile.

A variant of the Dijkstra's shortest path traversal can be used to calculate the LLCS coordinates of all RS's on the fly. The traversal is divided into two passes. In the first pass, the Dijkstra's algorithm starts from $RS_{base}^{to}$ and expands to the rest of the network, while ignoring all links connected to $RS_{base}^{from}$. During the traversal, the forward coordinates of all reached RS's are calculated or updated. Similarly, in the second pass, the Dijkstra's algorithm starts from $RS_{base}^{from}$ and ignores all links connected to $RS_{base}^{to}$, filling the backward coordinates of all reached RS's.

In the context of an LLCS, any link (including the base link) is related to the base link by the LLCS coordinates of its $RS^{from}$ and $RS^{to}$. Specifically, let a link to be related to the base link be called a ***test link*** and its end RS's be denoted as $RS_{test}^{from}$ and $RS_{test}^{to}$. Vector $v_{test} = \left[\boldsymbol{x}^{+}_{RS_{test}^{from}}, \boldsymbol{x}^{-}_{RS_{test}^{from}}, \boldsymbol{x}^{+}_{RS_{test}^{to}}, \boldsymbol{x}^{-}_{RS_{test}^{to}}\right]$ is defined as the relationship vector of the test link in the LLCS. With the relationship vector, the network distance between a crash $c_{base}$ on the base link and a crash $c_{test}$ on a test link can be easily calculated using Equations 3-7. Since the four coordinates in the relationship vector might result from different routings, there could be four possible crash-to-crash distances (Equations 4-7) whose geometric meanings are demonstrated in Figure 7. The final crash-to-crash distance should be the smallest possible distance. Besides the distance value, one can also tell if the two crashes were in the same traffic direction. For example, if the final distance is $d_F^+$ (upper right case in Figure 7), the centerline of the resulting route is bolded and the traffic directions of both crashes (green arrows) are on the same side of the centerline, meaning the two crashes (or links) were in the same traffic direction; otherwise, for $d_F^-$ and $d_T^+$, the two crashes were in the opposite traffic directions. Additionally, one can also determine whether $c_{test}$ happened upstream or downstream of $c_{base}$. For instance, $c_{test}$ happened upstream of $c_{base}$ if $d_T^+$ or $d_T^-$ is the final distance (when the test crash direction follows the bolded route); otherwise, $c_{test}$ happened downstream of $c_{base}$ (when the test crash direction departs from the bolded route).

$$d(c_{base}, c_{test}) = min(d_F^+, d_T^+, d_F^-, d_T^-) \tag{3}$$
$$d_F^+ = x^+_{RS_{test}^{from}} - os_{base} + os_{test} \tag{4}$$
$$d_T^+ = x^+_{RS_{test}^{to}} - os_{base} + (l_{test} - os_{test}) \tag{5}$$
$$d_F^- = x^-_{RS_{test}^{from}} + os_{base} + os_{test} \tag{6}$$
$$d_T^- = x^-_{RS_{test}^{to}} + os_{base} + (l_{test} - os_{test}) \tag{7}$$

where,

$d(c_{base}, c_{test})$= The network distance between $c_{base}$ and $c_{test}$, mile;
$d_F^+$= A possible distance via $RS_{test}^{from}$ forward coordinate, mile;
$d_T^+$= A possible distance via $RS_{test}^{to}$ forward coordinate, mile;
$d_F^-$= A possible distance via $RS_{test}^{from}$ backward coordinate, mile;
$d_T^-$= A possible distance via $RS_{test}^{to}$ backward coordinate, mile;
$x^+_{RS_{test}^{from}}$= Forward coordinate of $RS_{test}^{from}$, mile;

$x^-_{RS^{from}_{test}}$ = Backward coordinate of $RS^{from}_{test}$, mile;
$x^+_{RS^{to}_{test}}$ = Forward coordinate of $RS^{to}_{test}$, mile;
$x^-_{RS^{to}_{test}}$ = Backward coordinate of $RS^{to}_{test}$, mile;
$os_{base}$ = Link offset of $c_{base}$, mile;
$os_{test}$ = Link offset of $c_{test}$, mile;
$l_{test}$ = Length of the test link, mile.

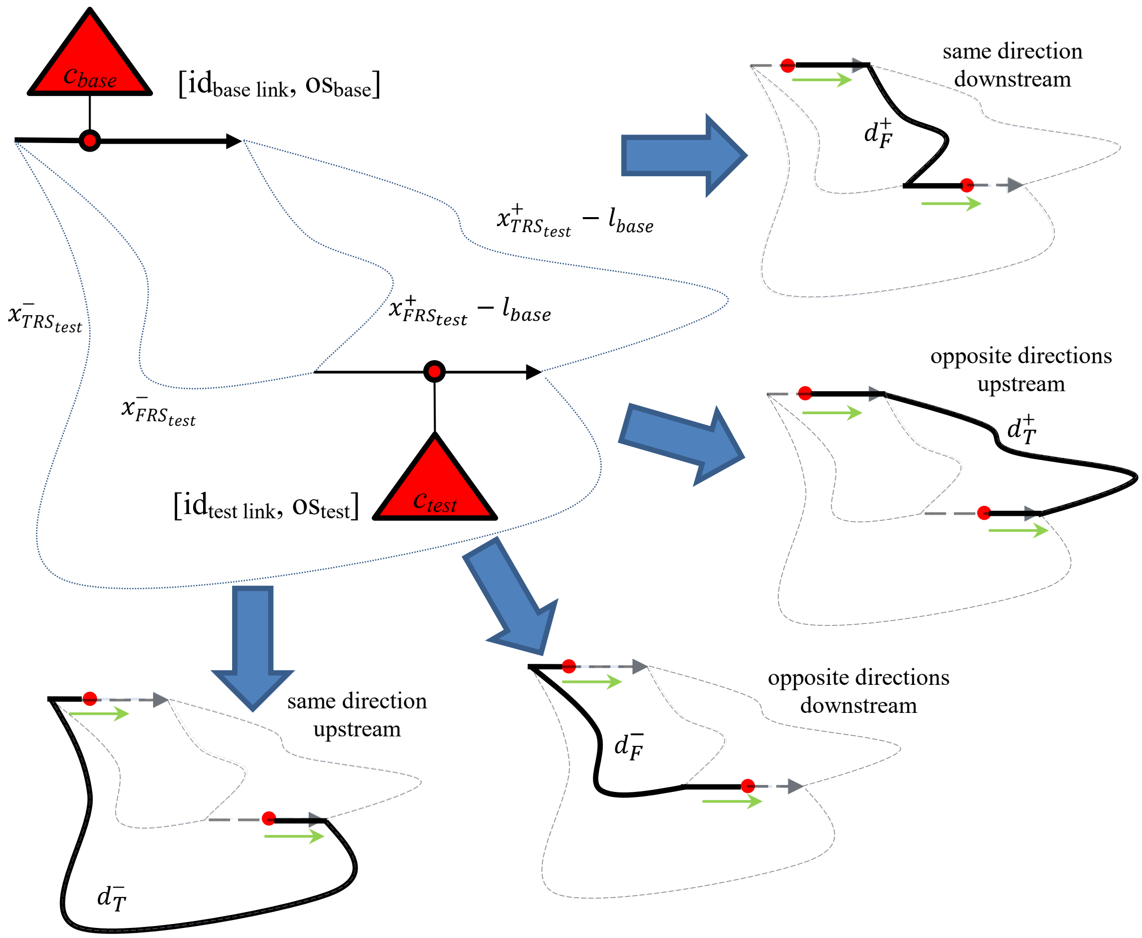



**FIGURE 7 Four possible distances between two crashes based on the relationship vector.**

### 3.3.2 Candidate Link

In the previous section, every link is assumed to be tested against the base link. However, given a particular spatial threshold of $D$ miles, a test link too far away from the base link is irrelevant to finding the near-by crash pairs. Only those links whose relationship vectors satisfy a certain condition may contain crashes within $D$ miles of the base link crashes. In fact, the condition is as simple as $\min\left(x^+_{RS^{from}_{test}} - l_{base}, x^-_{RS^{from}_{test}}, x^+_{RS^{to}_{test}} - l_{base}, x^-_{RS^{to}_{test}}\right) \le D$, where $l_{base}$

is the length of the base link. Links satisfying this condition are called candidate links and form a relatively small and relevant portion of the network (when $D$ is relatively small). The two passes of Dijkstra's traversal can stop expansion as early as any further RS to be reached has a forward coordinate larger than $l_{base} + D$ and a backward coordinate larger than $D$. Then, all links connected to the already reached RS's are all the candidate links.

### 3.3.3 The Algorithm

Figure 8 is the pseudo code of the proposed crash pairing algorithm. $L_{base}$ is assumed to be a preprocessed set of links containing at least one crash. The statement "find all candidate links" refers to the preparation of the relationship vectors for all candidate links in the LLCS as described in the above sections. $t(*)$ is the function of getting the time of a crash in minutes from a consistent time origin. $T$ and $D$ are the static thresholds in minutes and miles, respectively. It should be noted that the recorded time of crash could be slightly different from the time when the crash actually occurred. However, the authors do not expect it to have a significant impact on the results since a large time threshold of 5 hours was used. The statement "calculate $d(c_{base}, c_{cand})$" refers to Equations 3-7.

```
For each lk_base in L_base:
    Find all candidate links of lk_base as a set L_cand;
    For each candidate link lk_cand in L_cand:
        For each crash c_base in lk_base:
            For each crash c_cand in lk_cand:
                If 0 <= t(c_cand) – t(c_base) <= T:
                    Calculate d(c_base, c_cand);
                    If d(c_base, c_cand) <= D:
                        Add (c_base, c_cand) as a pair in the result;
```

**FIGURE 8 Pseudo code of crash pairing algorithm.**

Special cases should be treated differently. As illustrated in Figure 6, the longitudinal distance between crash B and crash C was only 0.15 miles and they occurred on opposite sides of the same highway. However, relying only on the network traversal of links, the resulting distance will go around $RS_{14}$ and be calculated as 0.25 miles. This type of unrealistic result is not desirable. In order to overcome this limitation, additional information from the STN was employed. An STN table of route-links was used to aid the links with their physical meanings. Each record in the route-link table tells which highway a link belongs to and in what direction. All links on the other side of the same highway are considered candidate links of the base link. When calculating the distance between a crash on the base link and a crash on the other side of the highway, the algorithm calculates the cumulative distances from the two crashes to a far upstream/downstream shared RS on the highway. The difference between these two cumulative distances is considered the distance between these two crashes. Additionally, when a shared RS could not be found, the algorithm further utilized another set of highway reference locations, reference points (RP). Each RP has its on-highway number, on-highway direction, RP number, and RP letter. If two RP's have the same on-highway number, RP number, and RP letter, they correspond to the same longitudinal position on the highway, even with different on-highway directions. Additionally, each RP, like a crash location, has a linear reference that maps it onto a link. Based on the above input, if two links on opposite sides of the same highway contains RP's with the same RP number and RP letter, there is a shared longitudinal position between them. Thus, instead of looking for a shared RS, the algorithm looks for a shared longitudinal position based on RP's.

### 3.3.4 Extension to Data Evolution

So far, the algorithm assumes that each link has not changed since its creation. Unfortunately, this assumption can be easily violated due to data evolution. In the highway network data, each link has a property, the valid period. More accurately speaking, the valid period of each link is represented by its DT_LINK_CURR field and LCM_DT_HSTL field. The DT_LINK_CURR field is the date when the link became operational in the highway network. The LCM_DT_HSTL field is the date when the link became historical. A link is only valid for the time period after the DT_LINK_CURR date and before the LCM_DT_HSTL date. For most of the links in the database, the LCM_DT_HSTL field is NULL, indicating that the link is still valid.

The valid periods of links introduce a challenge to the algorithm. Recall that when calculating a reference site's coordinate in a local linear coordinate system, the length of the shortest chain of links is used. Ignoring valid periods, the shortest chain of links is deterministic at any point in time. However, imagine in a chain of links, the first link was valid from January 1, 2009 to December 20, 2009 and the last link was valid from December 25, 2009 to present. This chain of links never physically existed, since it involves two links with valid periods that do not overlap. So, if valid periods of links are considered, a shortest chain of links might not exist or might be different at different points in time.

A versioning schema is introduced to address valid period problems. On top of links, a valid period is defined for a chain of links as the period when all of the links arevalid. If such a period does not exist, the chain of links is considered invalid and will not be used. The coordinate (forward or backward) of a reference site is no longer a single value, but instead a list of values. Each value is the version of coordinate for a specific time period. When computing the network distance between two crashes, only the coordinates that were valid when both of the two crashes happened are used. The introduction of the versioning schema has boosted the computation workload; but fortunately, due to all the speed improvements included in the algorithm, the final workload is more than acceptable.

### 3.3.5 Validation

The algorithm was implemented as a desktop application using the Java programming language. The current version of the application allows both offline and online running modes. In the offline mode, crash data and STN data should be downloaded onto the local file system beforehand. In the online mode when internet access is available, the application can connect to the WisTransPortal database and read all needed data from there. Mode choice is specified through a configuration file. Figure 9 illustrates the user interface of the application.

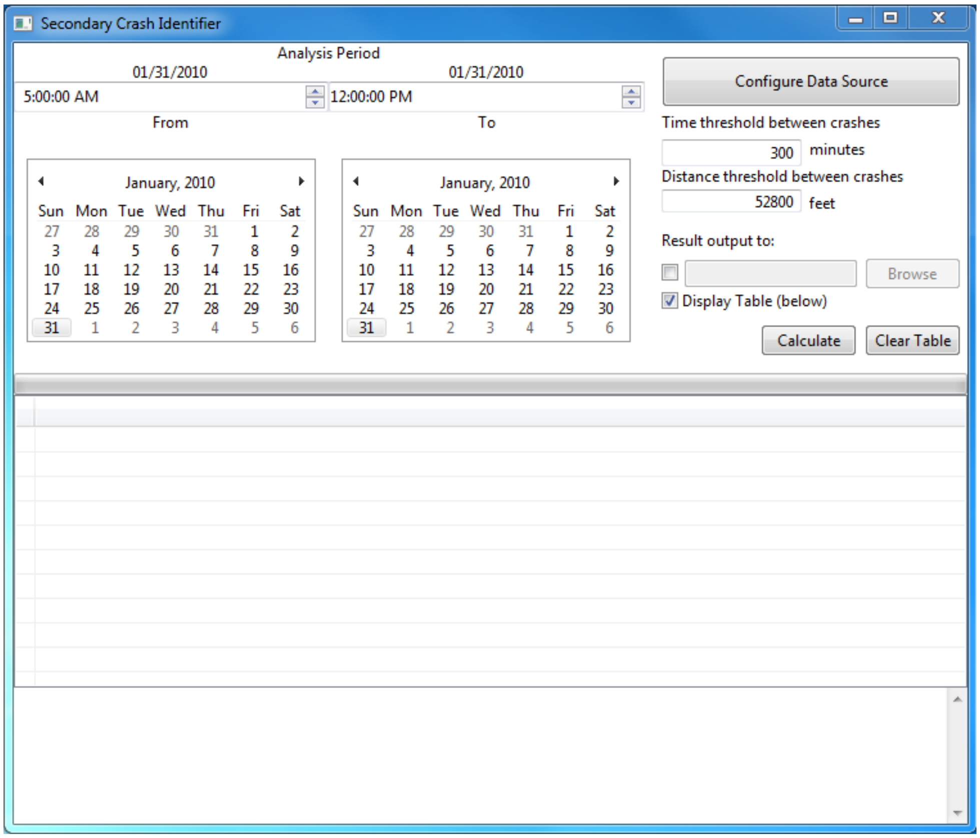



**FIGURE 9 User interface of the crash pairing program.**

The application passed several small independent tests (e.g., the entire stretch of a particular highway in the STN with crashes for several days) with manually extracted ground truths. In order to further validate the accuracy and the efficiency of the algorithm, a large scale network was tested. Since the ground truth in the large scale test was infeasible to be manually extracted, a relatively reliable ArcGIS based program was used as a mutual validation reference. The basic idea of the ArcGIS based program is to prepare a network dataset using the STN shapefile and the crash shapefile and use the buffer function of the NetworkAnalyst toolbox to find, for each crash, every other crash that is within a buffer network distance (the spatial threshold) from that crash. The ArcGIS based program was implemented in C++ using the ArcGIS APIs. Due to the unavailability of control over the buffer function of the

NetworkAnalyst toolbox, the ArcGIS based program was similar to the naïve algorithm of traversing the network for every pair of crashes, which provided the authors a chance to compare the efficiencies.

Both the pairing algorithm and the ArcGIS based program were tested on 10,922 crashes from the selected highway network of about 1,500 total miles in Wisconsin in 2010, with $D = 10$ miles and $T = 5$ hours. The pairing algorithm yielded 15,901 crash pairs, while the ArcGIS based program yielded 13,850 crash pairs. Of those crash pairs, 13,594 were identified by both systems. The ArcGIS based program captured 256 extra pairs, which were later found to be missed by the pairing algorithm due to computer precision problems, and did not hurt the validity of the pairing algorithm. The pairing algorithm captured 2,307 extra pairs, which were correct, but missed by the ArcGIS based program. In summary, the pairing program correctly identified more crash pairs than the ArcGIS based program. In addition, the ArcGIS based program finished the analysis in two and a half days, while the pairing algorithm finished in about 2 hours (30 times faster).

## 3.4 Stage 3: Pair-wise Filtering

For the purpose of identifying secondary crashes, the pairing algorithm produces an initial searching set, which without additional filtering, could be too vast to be useful. For example, in the year of 2010, the number of crash pairs fitting into the 5-hour-10-mile window is about 20,000, involving about 8,000 distinct crashes. Manually reviewing all these candidates is not only labor intensive but also unwise, because the 5-hour-10-mile window can include a considerable amount of false positive candidates (e.g., two crashes occurred at the same time while being 5 miles away from each other). Additionally, crashes on ramps are less likely to cause significant backup queues into the mainline freeways. Two filters are proposed below to select crash pairs that are more likely to have a primary-secondary relationship.

### 3.4.1 Ramp Filter

A candidate pair was excluded if its base crash happened on a ramp (on/off). The rationale behind this filter is that ramp crashes are rarely expected to cause secondary crashes on the mainline. To verify this, one year of crash pairs were sampled and manually reviewed. Among the 19,448 non-deer related and non-inclement weather related crash pairs in 2010 (5-hour-10-mile window), 797 involved an earlier crash happening on a ramp. A total of 85 samples were selected from the 797 crash pairs. One or two crash pairs were sampled from each of the 1-hour-5-mile intervals as shown in Figure 10. Manual review showed that none of the 85 samples contained a primary-secondary causal relationship. One of the sampled crash pairs involved two secondary crashes, but they were not related to each other; in addition, these two secondary crashes were captured by other crash pairs with their potential primary crashes.

The following procedure is used to identify if an earlier crash of a crash pair is a ramp crash. First, from the earlier crash's crash record, the link where the crash happened can be identified via the LINKID attribute. With the link ID, the chain(s) (another representation unit of the roadway for ArcGIS illustration) overlapping the link can be extracted from the DT_RDWY_LINK_CHN table maintained in the WisTransPortal database. Next, for each overlapping chain, there is an attribute named RDWY_CHN_RTE in the attribute table of the rdwy_chn coverage for ArcGIS. If the value of RDWY_CHN_RTE is "Ramp", then the corresponding chain is a ramp. If at least one of the overlapping chains is a ramp, the earlier crash is considered to have occurred on a ramp.

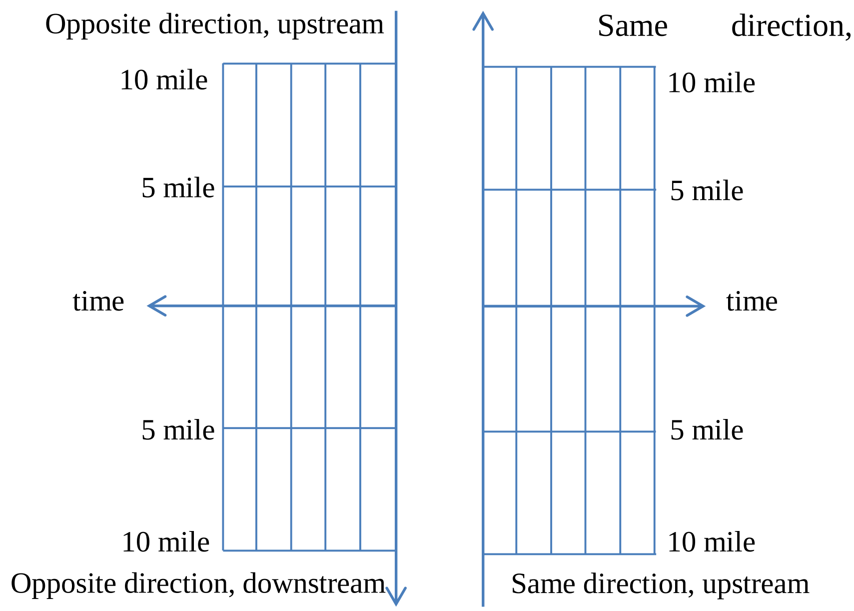



**FIGURE 10 Sampling intervals.**

### 3.4.2 Impact Area Filter

As mentioned in the literature review, crash pairs resulting from static thresholds could contain false primary-secondary pairs. These false pairs generally have unreasonable combinations of time distance and spatial distance. For example, a candidate pair whose time distance is 0 minutes but the spatial distance is 5 miles is certainly not a primary-secondary crash pair. Since secondary crashes have generally occur in or because of the queue caused by the primary incidents, queue theories have been commonly used to establish the time-varying impact area of the primary incidents to identify secondary crashes *(12–15, 17, 18, 21–24, 28, 45–47)*. Comparison of various queue estimation methods can be found in more general traffic research papers *(48, 49)*. Based on the literature review, none of the previous secondary crash studies used the shockwave model to estimate the impact area (IA) caused by a primary incident. In the current study, the IA of a crash is defined between two shockwaves: the queuing shockwave and the discharging shockwave. Mathematical representation for the evaluation of whether a crash occurred within the IA of another crash is given in Equations 8 through 10. Traffic flow for the prevailing condition ($q_1$) was obtained from the monthly average hourly traffic volume provided by the TRADAS detectors, with the same day of week and the same hour of day as the former crash. If the later crash happened outside the IA (in both directions) of the former crash, the crash pair will be excluded. On the other hand, secondary crashes could happen in the vicinity of the primary incident during its clearance. This type of secondary crash has typically been attributed to the "rubbernecking" effect *(20, 50)*. In order to capture these secondary crashes, a crash pair whose spatial distance (upstream or downstream in either traffic direction) was no larger than 1 mile and whose temporal distance was no larger than 1 hour was reserved, even if it did not satisfy the IA requirement.

$$a_2 \times (t - t_{clearance}) \leq d \leq a_1 \times t$$

8)

$$a_1 = (q_2 - q_1)/(k_2 - k_1)$$

9)

$$a_2 = (q_3 - q_2)/(k_3 - k_2)$$

10)

where,

$t$ The time between the former crash and the later crash, hour;

$t_{clearance}$ 1 hour (the simplified crash clearance time);

$d$ The network distance between the two crashes, mile;

$a_1$ The queuing shockwave speed, mile/hour;

$a_2$ The release shockwave speed, mile/hour;

$q_1$ The traffic flow of the prevailing condition, veh/hr/ln;

$k_1$ The density of the prevailing condition, veh/mile/ln. As a simplification, 65 mile/hr is assumed as the prevailing speed, and $k_1 = q_1/65$;

$q_2$ 0 veh/hr/ln (the traffic flow of the jam condition);

$k_2$ 352 veh/mile/ln (the density of the jam condition, assuming 15 feet minimum head to head distance between vehicles);

$q_3$ 1900 veh/hr/ln (the traffic flow of the saturated condition);

$k_3$ 1900/65veh/mile/ln (the density of the saturated condition).

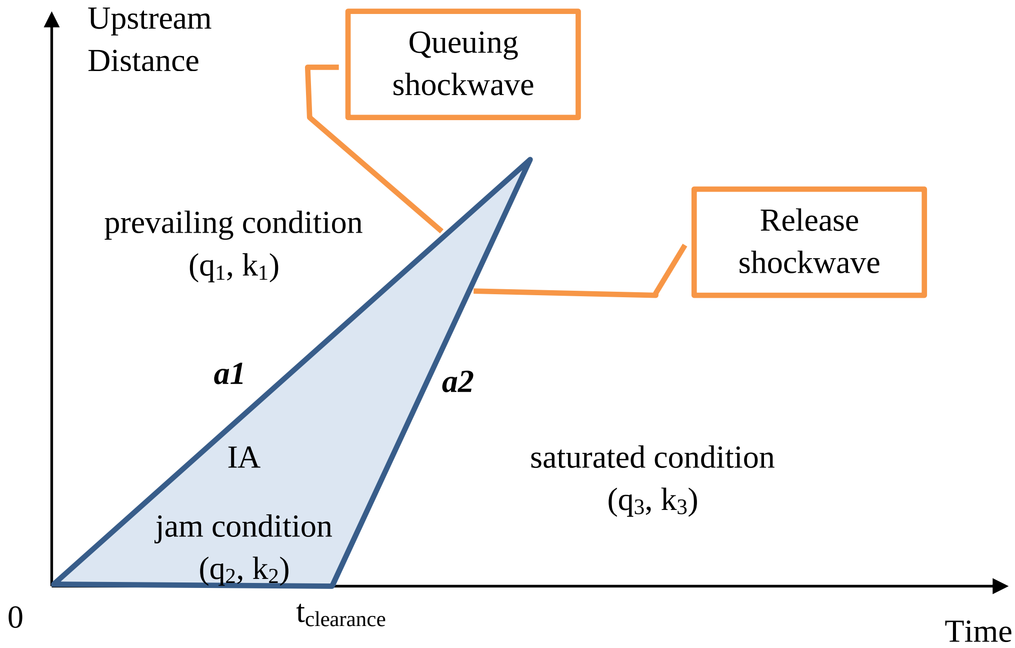



**FIGURE 11 Illustration of an impact area.**

## 3.5 Review of Crash Reports

After the 3-stage candidate selection, crash reports of all the remaining crash pairs were manually reviewed so actual secondary crashes could be identified and confirmed. At this stage, researchers reviewed the police reports of the earlier crash and the later crash side by side (Figure 12).

The later crash in the pair was reviewed first to judge if it was a secondary crash. Attention was paid in finding any statement in the police investigation that explicitly indicates a relation to any previous incident. Once the later crash was confirmed to be a secondary crash, the reviewer would review the first crash to see if it was the mentioned primary crash. When the crash identification number of the previous crash was mentioned in the later crash, the reviewer could simply compare it to that of the earlier crash in the pair. When such an identification number was not available, the reviewer would judge if the mentioned crash is the other crash in the crash pair by comparing accident features such as the number of vehicles involved. When the time difference between the two crashes in a pair is smaller than 15 minutes, the order of these two crashes might be unreliable since one crash could have happened before the other, while its notification time was later than the other's. In this case, both crashes should be reviewed as potential secondary crashes.

For each identified secondary crash and primary crash, several pieces of information were collected. First of all, the accident type of the crash must be noted. Normally, this information can be found in the field "Manner of Collision." However, this field is not always accurate. So, in addition to referring to the Manner of Collision field, the reviewer should also check the actual crash diagram and the crash narrative to verify what exactly happened. Common accident types showed in some preliminary results are rear-end, sideswipe same/opposite direction, angle, head-on, lost control, hit/hit by debris, struck median, run-off, vehicle failure, etc. When a crash involved multiple accident types, the notation should list all types involved with a comma for separation, ordered alphabetically by first letter. Second, the location type and the traffic conditions should be described. Common location types include divided/undivided highway segment, on ramp, off ramp, ramp terminal, signal intersection, stop control intersection, yield control intersection, etc. Traffic condition can be normal traffic, slow traffic, and stopped traffic. Third, any special roadway conditions should be mentioned. Construction zone, lane closure, and the awareness of another accident are of particular interest and should also be flagged when mentioned. Fourth, the number of involved lanes should be recorded to provide a sense of the direct impact that a crash had on the traffic flow.

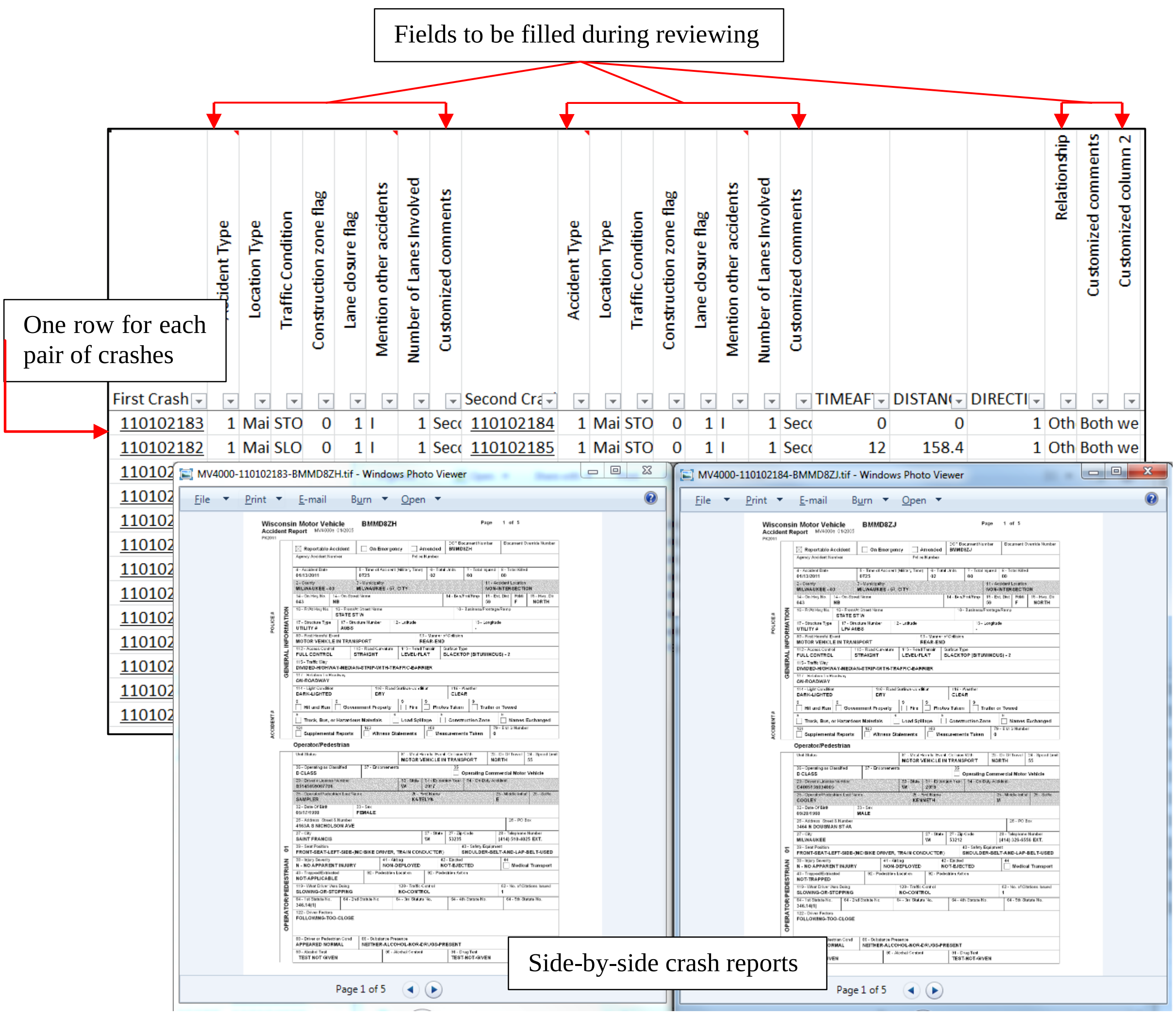



**FIGURE 12 Illustration of the manual review environment.**

# CHAPTER 4 SECONDARY CRASH IDENTIFICATION RESULTS

Secondary crash identification was conducted annually from 2007 to 2011 using the method described in the previous section. A temporal-spatial window of 5-hours-10-miles was chosen for the time-space pairing stage. This window was suggested by the project panel based on their experience in urban areas such as Milwaukee. For each year, initial input crash data included all crashes in that year, as well as crashes that happened within the last five hours of the previous year and the first five hours of the next year. This is corresponding to the 5-hour temporal threshold so any crash pair crossing the new year's boundary could be caught. Figure 13 summarizes the step-wise filtering results for finding the primary-secondary candidate pairs. The numbers of crashes and crash pairs that remained after each filtering step are highlighted in green, while those excluded are grayed out. After stage 1, the candidate crashes were reduced to 4.9% - 5.8% of the initial input crashes (about 94.2% - 95.5% in reduction rate). After stage 2, the candidate crashes were reduced to 2.5% - 3.5% of the initial input crashes (about 96.5% - 97.5% in reduction rate). After stage 3, the initial crashes were reduced to about 0.8% - 1.1% (98.9% - 99.2% in reduction rate). The resulting numbers of crash pairs ranged from 661 to 1,008 for each year, which were more amenable for manual review in a relatively short time period.

| | | | | Stage 1: Initial Filtering | | | | Stage 3: Pair-wise Filtering | |
|---|---|---|---|---|---|---|---|---|---|
| | | | *Entire database** | On selected STN highways | Non-inclement weather related | Non-deer related | *Stage 2: Time-Space Pairing* | First crash not on ramp | Second crash in IA or unsure** |
| **2007** | *Included* | *# of crashes* | 137961 | 14671 | 8910 | 7054 | 4106 | 3902 | 1222 |
| | | *# of pairs* | | | | | 7897 | 7130 | 957 |
| | *excluded* | *# of crashes* | | 123290 | 5761 | 1856 | 2948 | 916 | 3693 |
| | | *# of pairs* | | | | | | 767 | 6173 |
| **2008** | *Included* | *# of crashes* | 137923 | 15490 | 7836 | 6220 | 3444 | 3219 | 1076 |
| | | *# of pairs* | | | | | 6501 | 5855 | 851 |
| | *excluded* | *# of crashes* | | 122433 | 7654 | 1616 | 2776 | 839 | 3038 |
| | | *# of pairs* | | | | | | 646 | 5004 |
| **2009** | *Included* | *# of crashes* | 121942 | 12495 | 7709 | 5977 | 3164 | 2943 | 924 |
| | | *# of pairs* | | | | | 5463 | 4843 | 661 |
| | *excluded* | *# of crashes* | | 109447 | 4786 | 1732 | 2813 | 804 | 2766 |
| | | *# of pairs* | | | | | | 620 | 4182 |
| **2010** | *Included* | *# of crashes* | 120851 | 12513 | 8732 | 7034 | 4231 | 4029 | 1342 |
| | | *# of pairs* | | | | | 8665 | 7868 | 1008 |
| | *excluded* | *# of crashes* | | 108338 | 3781 | 1698 | 2803 | 1005 | 3838 |
| | | *# of pairs* | | | | | | 797 | 6860 |
| **2011** | *Included* | *# of crashes* | 125106 | 13829 | 8569 | 6734 | 3859 | 3578 | 1111 |
| | | *# of pairs* | | | | | 7092 | 6271 | 819 |
| | *excluded* | *# of crashes* | | 111277 | 5260 | 1835 | 2875 | 1000 | 3387 |
| | | *# of pairs* | | | | | | 821 | 5452 |

* For each year, the count of crashes in the entire database include 5 hours into the previous year and the next year.
** The uncertainty arised when traffic information is not available for shock wave analyses

**FIGURE 13 Filtering results of primary-secondary candidate pairs.**

Candidate primary-secondary crash pairs at the end of the 3rd stage were then manually reviewed for secondary crash confirmation. A candidate pair might involve two independent crashes, which would be excluded. A candidate pair could also only involve one or two secondary crashes, without the primary crash being captured. There were also a considerable amount of pairs that captured both the primary crash and the secondary crash. Table 3 summarizes the review results. For each year, only about half of the identified secondary crashes were found with their primary crashes. A number of reasons were found for the missing primary crashes, which include undocumented primary crashes and primary crashes not coded on an STN link, among other data quality issues.

**TABLE 3 Manual Review Results**

| Year | 2007 | 2008 | 2009 | 2010 | 2011 |
|---|---|---|---|---|---|
| **# of pairs involving secondary crashes** | 82 | 63 | 69 | 73 | 84 |
| **# of pairs involving both the primary and the secondary crashes** | 44 | 41 | 37 | 67 | 40 |
| **# of distinct primary crashes** | 39 | 39 | 37 | 65 | 35 |
| **# of distinct secondary crashes** | 62 | 53 | 51 | 73 | 63 |

# CHAPTER 5 ANALYSIS

Analyses were performed on the identified secondary crashes, as well as on their relationship to the primary crashes. For secondary crashes, the analyses were focused on revealing their temporal and spatial distributions, as well as identifiable contributing factors. For primary and secondary crash pairs, the analyses were focused on summarizing the characteristics of the time lapse and the spatial relation between a primary crash and a secondary crash. Table 4 gives an overview of secondary crashes by their relative traffic direction to that of the primary crash and by manner of collision. Most secondary crashes happened on the same direction of the roadway as the primary crash. Over 75% of secondary crashes are rear-end crashes; sideswipes in the same direction are the second major collision type of secondary crashes; other crash types include sideswipes from the opposite direction, angle crashes, collisions with infrastructure, and run-off-road crashes.

**TABLE 4 Overview of Identified Secondary Crashes**

| Year | 2007 | 2008 | 2009 | 2010 | 2011 |
|---|---|---|---|---|---|
| **Number of confirmed secondary crashes** | 62 | 53 | 51 | 73 | 63 |
| **Distribution of directions (same/different directions*)** | 38/6 | 34/7 | 30/7 | 52/15 | 36/4 |
| **Number (percentage) of rear-ends** | 48 (77.4%) | 42 (79.2%) | 44 (86.3%) | 50 (68.5%) | 48 (76.2%) |
| **Number (percentage) of sideswipes in the same direction** | 5 (8.1%) | 8 (15.1%) | 4 (7.8%) | 11 (15.1%) | 5 (7.9%) |
| **Number (percentage) of other types of crashes** | 9 (14.5%) | 3 (5.7%) | 3 (5.9%) | 12 (16.3%) | 10 (15.9%) |

* The two numbers do not add up to the number of confirmed secondary crashes, because some secondary crashes were identified without their primary crashes.

## 5.1 Temporal Trends of Secondary Crashes

Temporal distributions of the identified secondary crashes were compared to general crashes on the same roadways. General crashes were those crashes remaining after Stage 1 in Figure 13, excluding crashes on ramps. Temporal distributions were calculated on three scales: 1) month, 2) day of week, and 3) time of day, and are illustrated in Figure 14, Figure 15, and Figure 16, respectively. Secondary crashes and general crashes have similar trends for day of week but different trends for month and for time of day. More secondary crashes (by percentage) than general crashes happened in late night hours and between November and January; a lower percentage of secondary crashed compared to general crashes occurred in the May-June period.

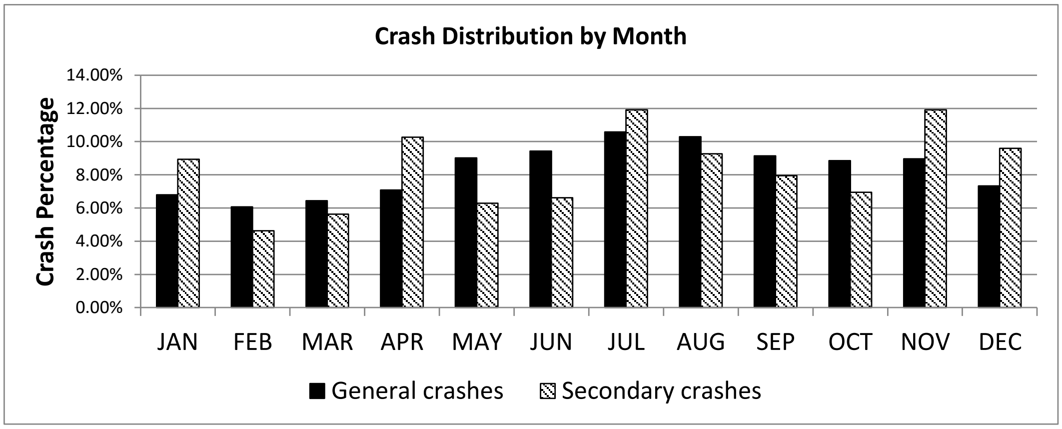



**FIGURE 14 Crash distribution by month.**

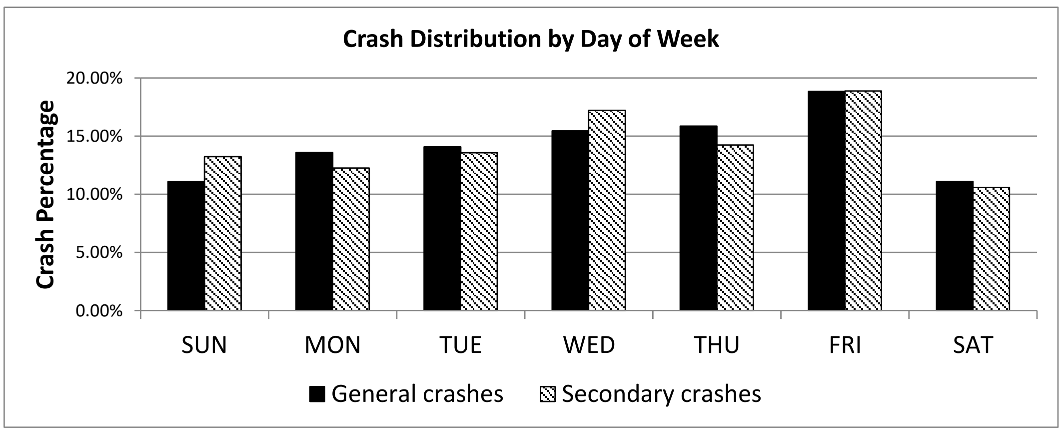



**FIGURE 15 Crash distribution by day of week.**

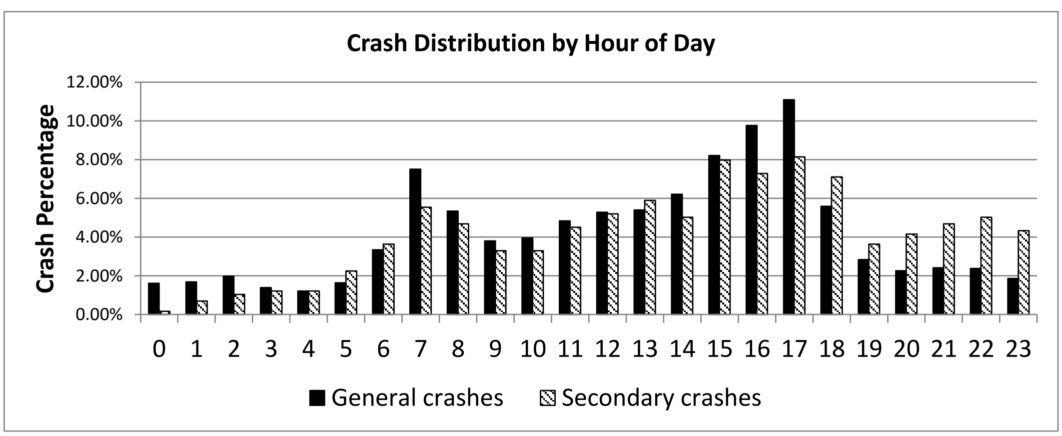



**FIGURE 16 Crash distribution by hour of day.**

## 5.2 Secondary Crash Hotspots

In this project, the spatial distribution of secondary crash frequency was computed to identify hotspots. The studied highways were split into 1-mile segments (with a small number of less-than-1-mile segments at the road ends and connectors) and counts of secondary crashes on each of those segments were obtained. Figure 17 shows the histogram of secondary crash frequencies. For either or both traffic directions, the majority of segments contain no secondary crashes (secondary crash frequency being 0). The next common frequency is 1 secondary crash in five years, followed by 2 secondary crashes in five years. A frequency of 3 secondary crashes or above becomes rare. Thus, the secondary crash frequency is divided into four categories: 0, 1, 2, and 3+. A frequency of 3+ indicates a hotspot.

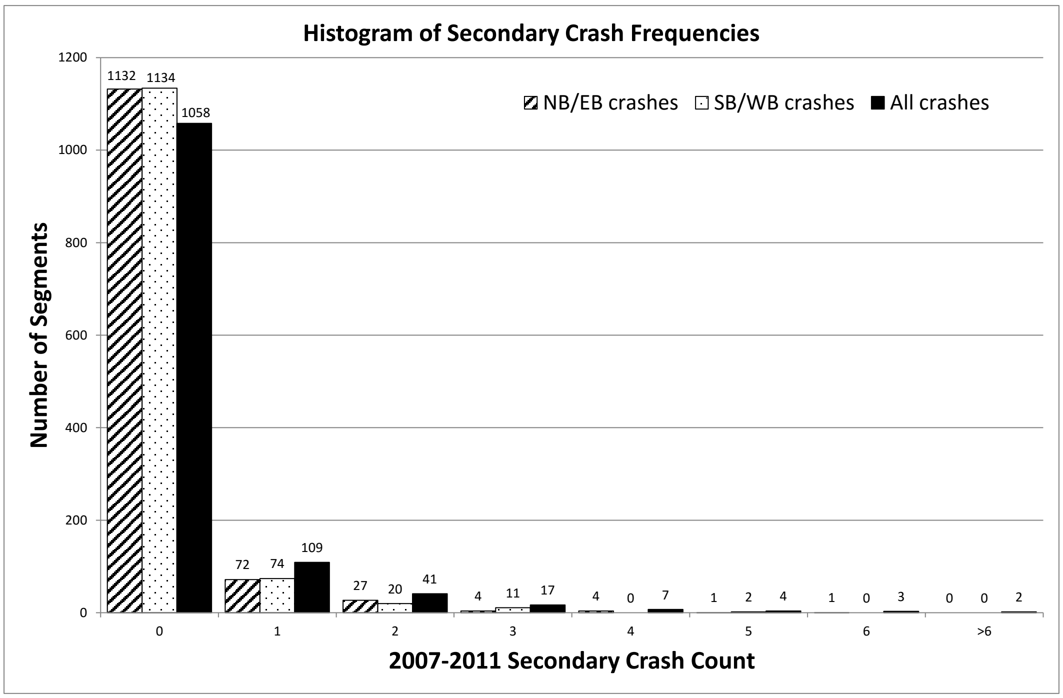



**FIGURE 17 Histogram of secondary crash frequencies (2007 – 2011).**

Figure 18 visualizes the spatial distribution of secondary crash frequencies at the state level. Segments with higher secondary crash intensities are displayed with thicker lines. Visual clusters clearly reveal hotspots in Milwaukee, Madison, Green Bay, and Appleton. Zoomed-in snapshots of secondary crash intensities by traffic direction are illustrated in Figure 19 to Figure 22 for these city areas. Instead of line weights, line colors are used to distinguish different secondary crash intensities at the city level, allowing better observation of underlying roadway layouts. For all four cities, the hotspot locations are different between the NB/EB traffic directions and the SB/WB traffic directions, but are all within a 1-mile range from an interchange. Interchanges are well known bottleneck locations along highways. Heavy traffic at these bottleneck locations can easily lead to backup queues when an incident happens, resulting in higher risks for secondary crashes. Observation of secondary crash hotspots around interchanges complies with such expectations.

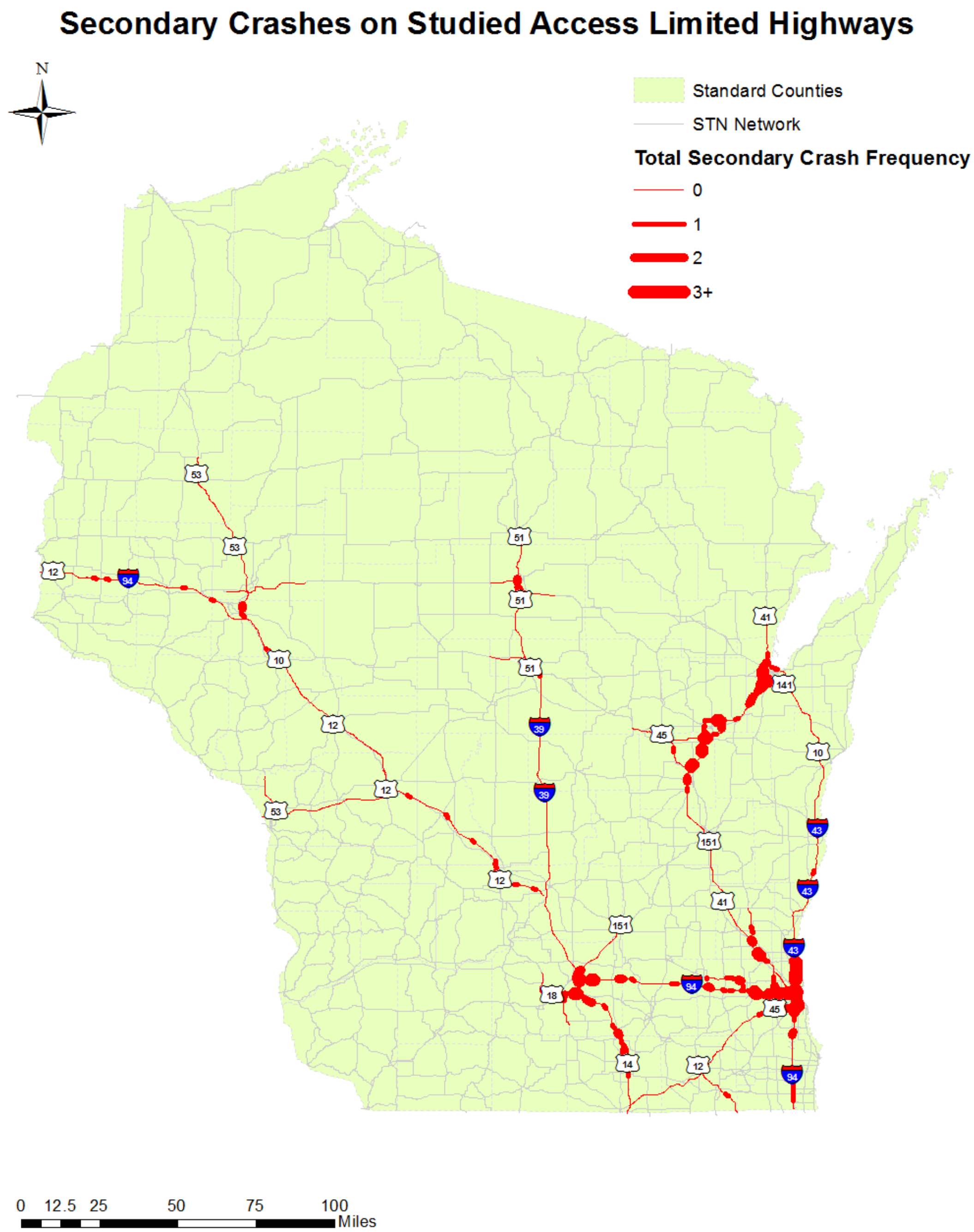



**FIGURE 18 Statewide overview of secondary crash frequencies.**

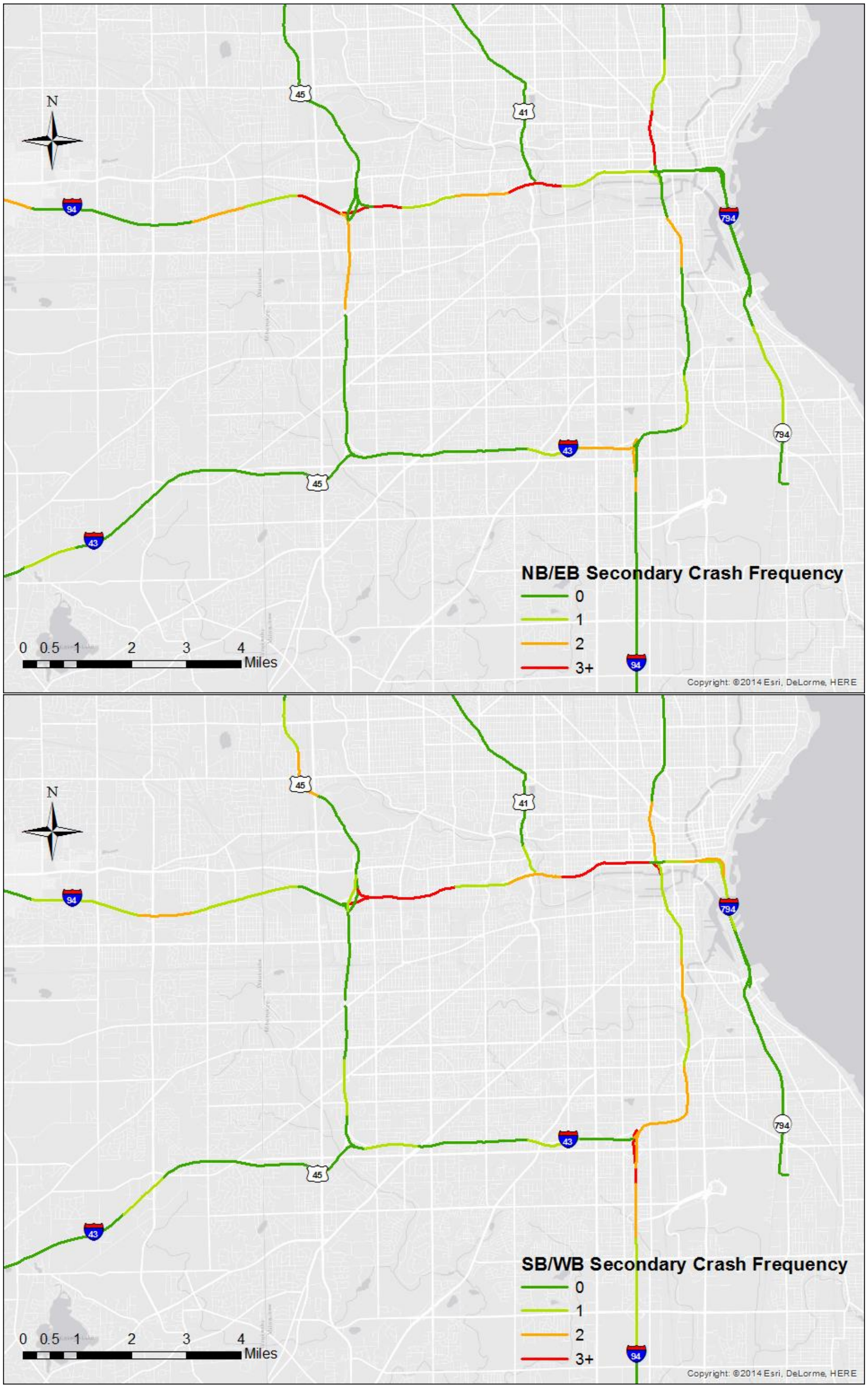



**FIGURE 19 Secondary crash frequency in Milwaukee (2007-2011).**


**FIGURE 20 Secondary crash frequency in Madison (2007-2011).**

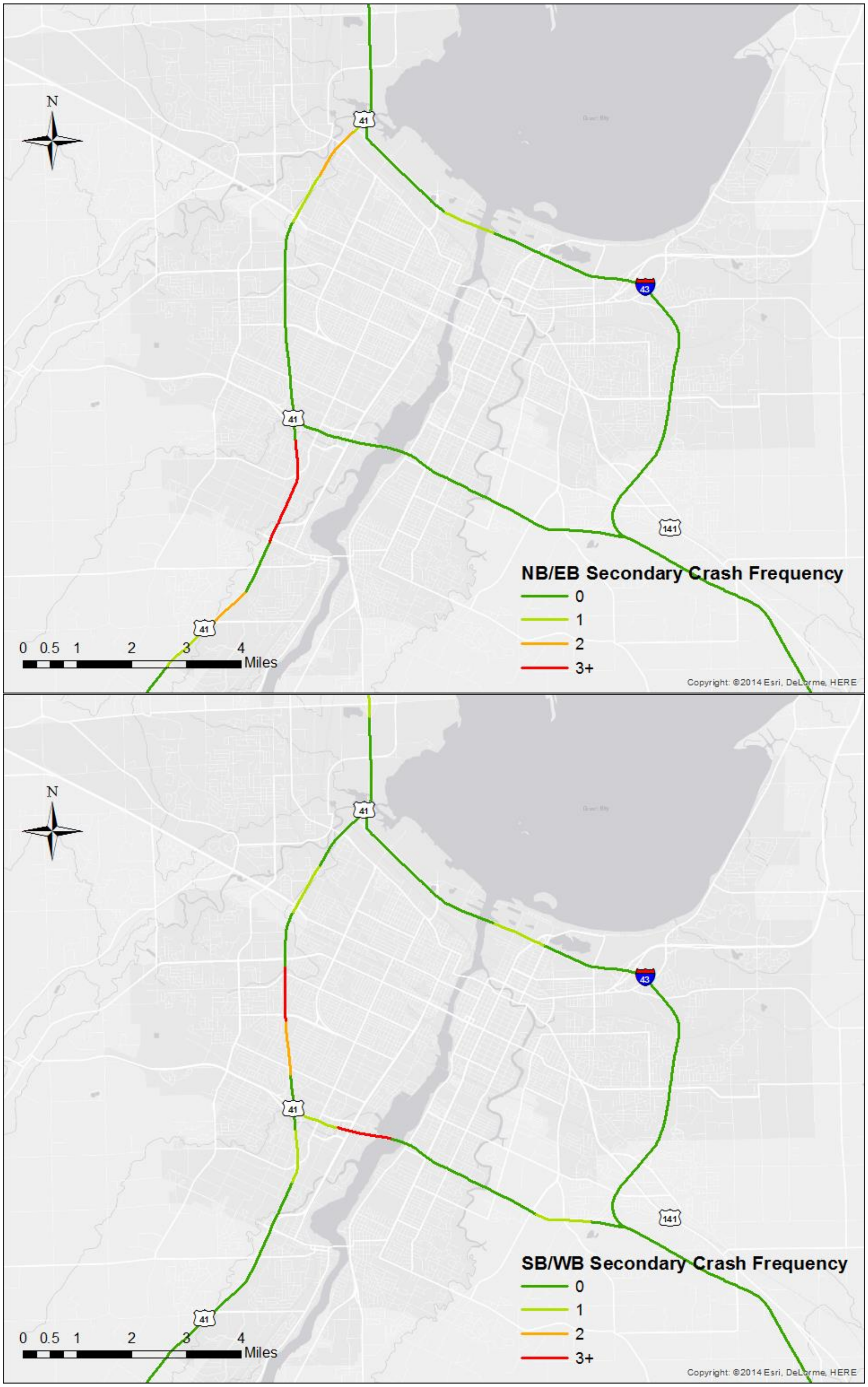
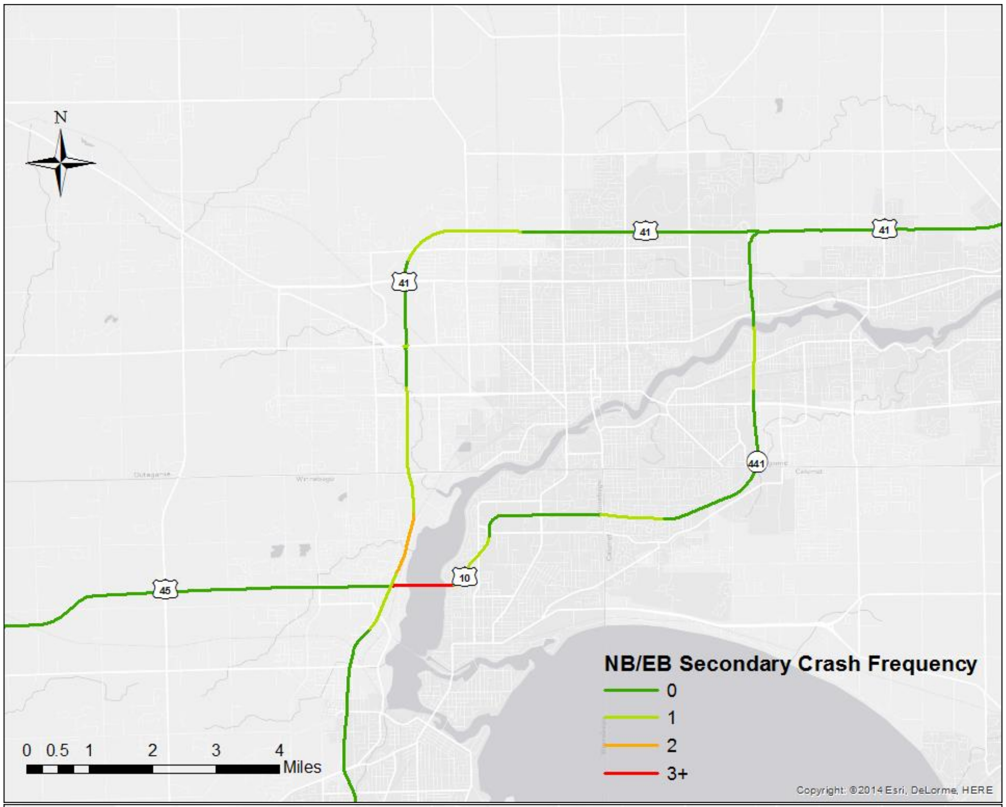



**FIGURE 21 Secondary crash frequency in Green Bay (2007-2011).**



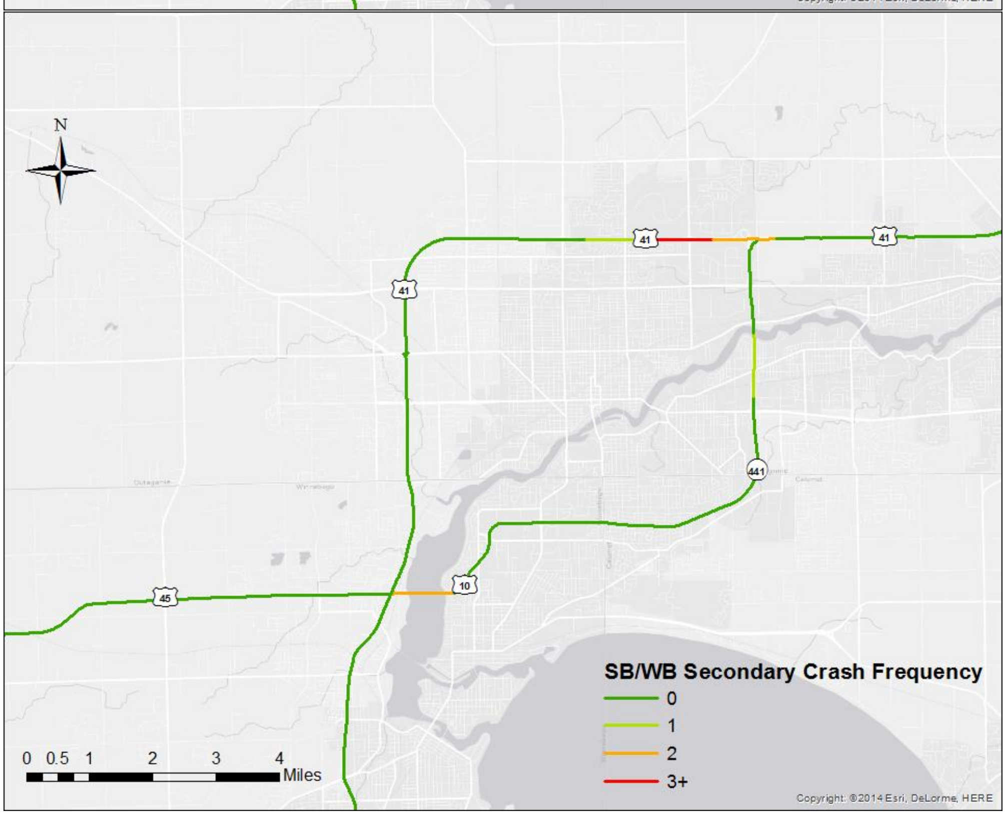




**FIGURE 22 Secondary crash frequency in Appleton (2007-2011).**

## 5.3 Contributing Factors to Secondary Crashes

Figure 23 and Figure 24 show secondary crash frequencies by highway factors and driver factors, respectively. An identified secondary crash might have none, one, or more than one of these factors labeled. Thus, the illustrated frequencies do not add up to the total number of identified secondary crashes. In Figure 23, inattentive driving contributes to the most secondary crashes, closely followed by debris prior to accident. While debris normally appeared near primary crash scenes, inattentive driving conditions could be present either in the vicinity of the primary crash or in stop-and-go traffic. Construction zone is the third influential highway factor, which normally involves closing traffic lanes. Obscured visibility could be caused by foggy weather, in which a series of related crashes have happened and been observed by the report reviewers. From the drivers' perspective, Figure 24 indicates that following too closely, inattentive driving, failure to control vehicle, and too fast for conditions are the four leading factors of secondary crashes. These factors are typically related to rear-end crashes, which is the major type of secondary crash, as noted in Table 4.

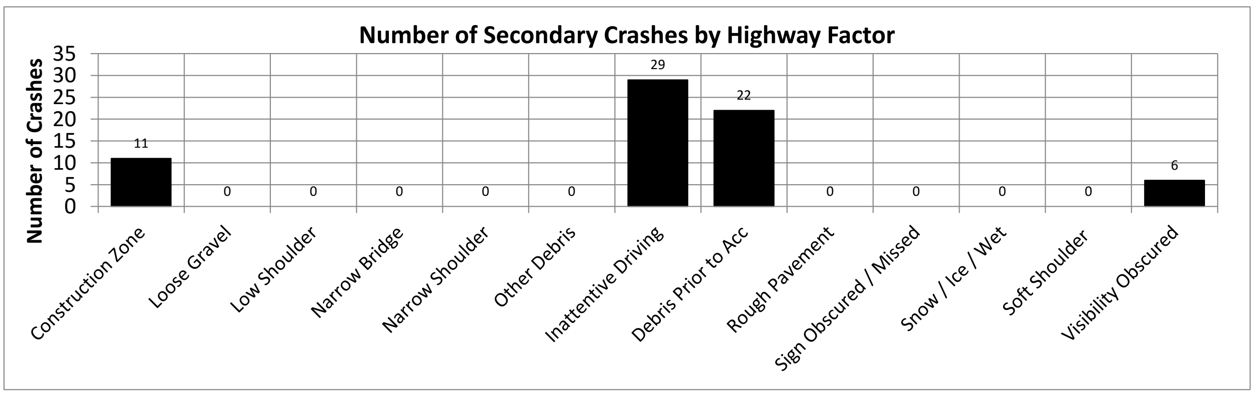



**FIGURE 23 Secondary crash frequency by highway factor.**

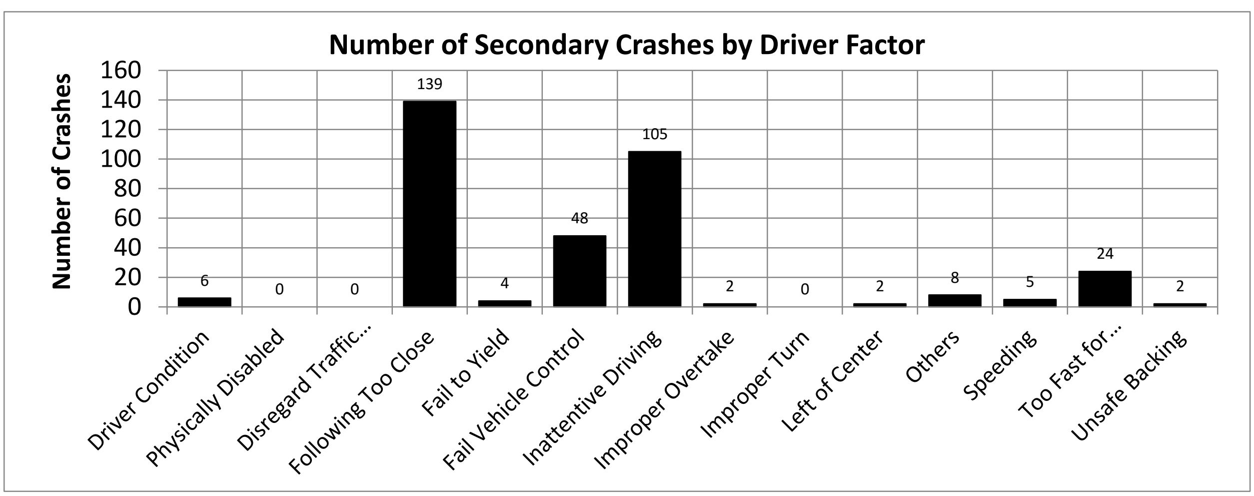



**FIGURE 24 Secondary crash frequency by driver factor.**

## 5.4 Spatial and Temporal Relationship between Primary Crashes and Secondary Crashes

Figure 25 shows the distribution of the time after primary crashes. Over 67% secondary crashes (whose primary crashes were identified) occurred within 20 minutes of the primary crash. Almost all secondary crashes happened within one hour of the primary crash. The average time of a secondary crash after a primary crash was around 17 minutes.

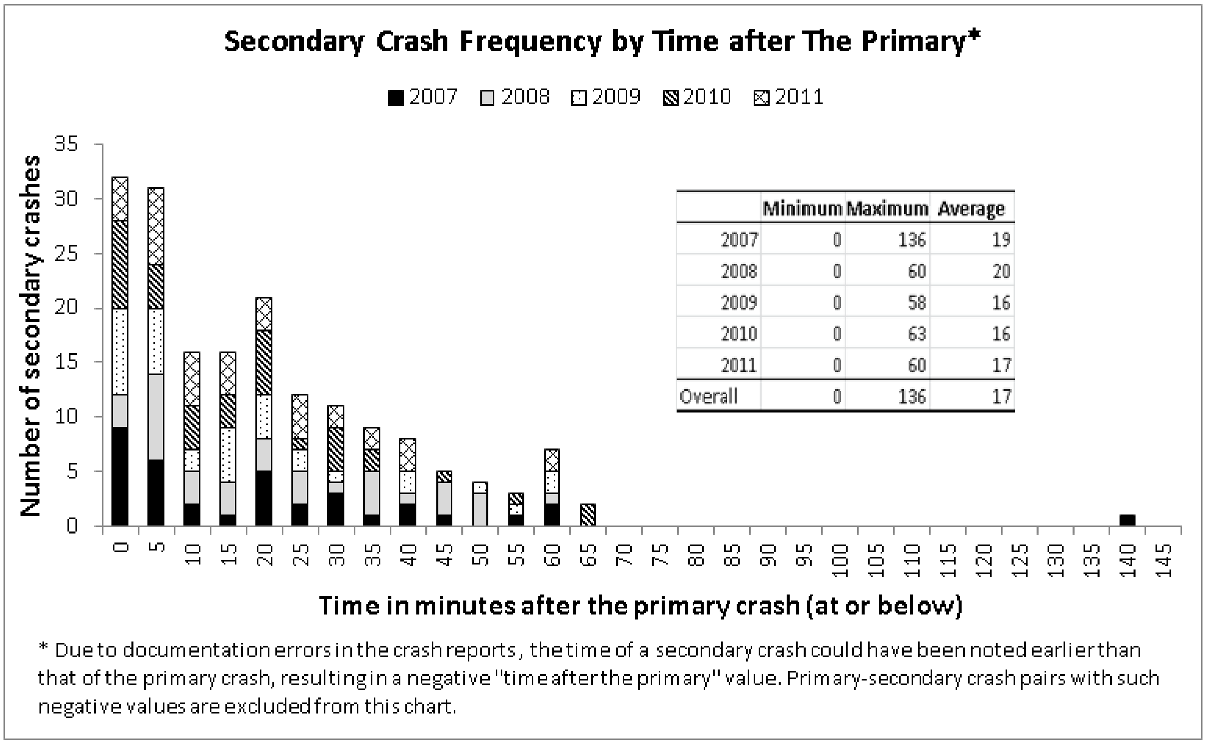



**FIGURE 25 Time distance distribution.**

Figure 26 shows the statistics of the distance between a secondary crash and its primary crash in different positions. The average distance for a secondary crash located upstream in the same traffic direction as the primary crash is 0.29 miles. For upstream in the opposite traffic direction, the average is 0.37 miles. For downstream in both directions, the averages are around 0.15 miles. The average upstream distances are longer than the downstream distances. According to observations from crash reports, downstream secondary crashes were primarily caused by rubbernecking, slow traffic passing the primary crash scene, or debris from the primary crash. Rubbernecking and debris were also common reasons for secondary crashes happening on the opposite side of traffic; the primary crash being a cross-median crash was also another reason. Upstream secondary crashes were majorly caused by congestion from the primary crash, while rubbernecking and debris could also cause near upstream secondary crashes.

| Upstream | | Downstream | |
|---|---|---|---|
| Minimum | 0.01 | Minimum | 0.01 |
| Maximum | 1.00 | Maximum | 0.74 |
| Average | 0.37 | Average | 0.16 |
| ***Opposite Direction*** | | ***Same Direction*** | |
| Minimum | 0.00 | Minimum | 0.00 |
| Maximum | 0.37 | Maximum | 1.94 |
| Average | 0.14 | Average | 0.29 |
| **Downstream** | | **Upstream** | |

**FIGURE 26 Network distance characteristics.**

# CHAPTER 6 CONCLUSIONS

An efficient and effective procedure was developed for identifying secondary crashes from statewide historical data. This procedure is two-fold. In the first part, three-stage filtering is applied to generate a list of potential primary-secondary crash pairs. In the second part, crash reports are manually reviewed to validate actual secondary crashes from the potential pairs. Efficiency is gained through the first part of the procedure, which is fully automated in a custom java program. Utilizing a linear referencing system for crash pairing in the program is a major advantage, making the program much faster and more accurate than an ArcGIS based counterpart developed in this research.

A total of 302 secondary crashes were identified through a 5-year time span on about 1,500 miles of access controlled highways in Wisconsin. More than half of these secondary crashes were found with their primary crashes. Missing primary crashes were the result of limited data such as undocumented crashes or crashes missing linear referencing information. Rear-end crashes compose 76% of the 302 secondary crashes. Secondary crashes happening in the same traffic direction of the primary crash are about twice as common as those in the opposite direction. Secondary crashes and general crashes have similar trends for day of week. However, more secondary crashes (by percentage) than general crashes happened in late night hours and in the November-January period; fewer occurred in May-June period. Secondary crash hotspots were identified based on secondary crash counts for each segment in the five year period. Segments with 3 or more secondary crashes were considered hotspots, normally located within 1-mile from an interchange. Inattentive driving conditions, debris, construction zones, and visibility obstructions (e.g., fog) are leading highway factors to secondary crashes. Freeway service patrols are advised to increase patrol intensity in the suggested hotspot range and under the above highway conditions. Following too closely, inattentive driving, failure to control vehicle and speeding are leading factors for secondary crashes from the driver side. The average time lapse between a primary crash and a secondary crash is 17 minutes. The average distances from an upstream secondary crash to the primary crash are 0.29 miles and 0.37 miles for the same side of traffic and the opposite side of traffic, respectively; the average distances from a downstream secondary crash to the primary crash are around 0.15 miles for both traffic directions. The longer upstream distances imply the crucial role of queuing as a cause of secondary crashes. Fast incident clearance and traffic restoration should receive no less emphasis based on the current finding. Rubbernecking and debris were two common reasons for secondary crashes that happened in the opposite direction or downstream of the primary crash. Cross-median primary crashes could also cause secondary crashes on the opposite side of the initial crash. Congestion or slow traffic caused by primary crashes contributed to both upstream secondary crashes, as well as downstream secondary crashes within a short distance from the primary crash scene. Police officers are recommended to extend their investigation area and increase attention for secondary crashes when traffic is congested.

In spite of the efficiency in narrowing down the search space for secondary crashes, the proposed procedures are still limited by data quality. The assumption of the completeness of the crash database can be violated and some primary crashes and secondary crashes were missing. Also, since the pairing algorithm relies on the link based linear referencing system, when the referencing information (i.e, link ID and offset) was missing from a crash record, the algorithm would lose track of that crash. In addition to crash data, there was also limitation in the traffic data. During the IA filter step, real time traffic volumes should be used. However, due to the limitations of detector coverage, the traffic volumes could not be retrieved for all potential primary crashes at their occurrences. When the traffic volume at the desired location was missing, the nearest upstream or downstream volume was used; similarly, when the traffic volume was not available for the primary crash hour, the volume in the nearest earlier hour was used. The unavailability or poor quality of crash reports were also a challenge to the study. Without the police narrative in the report, a true secondary crash could not be confirmed. In addition to the above limitations, the primary incidents in this study were constrained within crashes and need to be further extended to all other incident types. Also, the simple shockwave model for IA estimation can be refined with high resolution traffic data.

Three major future works are recommended. First, to make the whole workflow of secondary crash identification more automatic, optical character recognition (OCR) and artificial intelligence (AI) might be employed to assist researchers in manually reviewing police reports. Second, more years of data can be collected to establish a larger sample of secondary crashes for more comprehensive statistical analyses. Last but not least, crashes in inclement weather were not included in this analysis because the objective was to analyze secondary crashes that can be mitigated by TIM strategies. The authors realize that secondary crashes occur in inclement weather and recommend that future studies should examine the impact of weather on secondary crashes.

For future crash documentation, the authors recommend that two dedicated fields being added to the crash report form. First, a field indicating the traffic condition is recommended. This field can be a multiple selection set where police officers can check one among several levels of traffic intensity, such as free flow, moderate, slow, stop-and-go, stopped, etc. Another suggested field is the "accident number of a related crash." When no related crash was found within the vicinity of the current crash, this field should be left blank; when one or more related crashes were found, this field should be noted with the accident number of the crash closest in time, either before or after; when the accident number cannot be retrieved, "unknown crash" should be used. Police officers are advised to increase their attentions for the area within a decent distance up- and down-stream of the current crash location in both traffic directions as they fill out the two additional fields, especially when traffic is congested. Addition and proper filing of these two fields will facilitate future secondary crash identification and analysis.

## REFERENCES


1. Schrank, D., B. Eisele, and T. Lomax. *TTI's 2012 Urban Mobility Report*. 2012.
2. Margiotta, R., R. Dowling, and J. Paracha. *Analysis, Modeling, and Simulation for Traffic Incident Management Applications*. Washington, D.C., 2012, p. 54.
3. Khattak, A. J., X. Wang, and H. Zhang. Are Incident Durations and Secondary Incidents Interdependent? *Transportation Research Record: Journal of the Transportation Research Board*, No. 2099, 2009, pp. 39–49.
4. Karlaftis, M. G., S. P. Latoski, N. J. Richards, and K. C. Sinha. ITS Impacts on Safety and Traffic Management: An Investigation of Secondary Crash Causes. *Journal of Intelligent Transportation Systems*, Vol. 5, No. 1, 1999, pp. 39–52.
5. Moore, James E., I., G. Giuliano, and S. Cho. Secondary Accident Rates on Los Angeles Freeways. *Journal of Transportation Engineering*, Vol. 130, No. 3, 2004, pp. 280–285.
6. Owens, N., A. Armstrong, P. Sullivan, C. Mitchell, D. Newton, R. Brewster, and T. Trego. *Traffic Incident Management Handbook*. Washington, D.C., 2010, pp. 1–116.
7. Dunn, W. M., and S. P. Latoski. *NCHRP Synthesis of Highway Practice 318: Safe and Quick Clearance of Traffic Incidents*. Washington, D.C., 2003, pp. 1–143.
8. Green, E. R., J. G. Pigman, J. R. Walton, and S. Mccormack. Identification of Secondary Crashes and Recommended Countermeasures to Ensure More Accurate Documentation. 2012.
9. Raub, R. Occurrence of secondary crashes on urban arterial roadways. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1581, 1997, pp. 53–58.
10. Hirunyanitiwattana, W., and S. P. Mattingly. Identifying Secondary Crash Characteristics for California Highway System. 2006.
11. Zhan, C., L. Shen, M. A. Hadi, and A. Gan. Understanding the Characteristics of Secondary Crashes on Freeways. 2008.
12. Zhan, C., A. Gan, and M. Hadi. Identifying Secondary Crashes and Their Contributing Factors. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2102, Dec. 2009, pp. 68–75.
13. Chilukuri, V., and C. Sun. Use of Dynamic Incident Progression Curve for Classifying Secondary Accidents. 2006.
14. Orfanou, F. P., E. I. Vlahogianni, and M. G. Karlaftis. Detecting Secondary Accidents in Freeways. 2011.
15. Chou, C.-S., and E. Miller-Hooks. Simulation-Based Secondary Incident Filtering Method. *Journal of Transportation Engineering*, Vol. 136, No. 8, 2010, pp. 746–754.
16. Haghani, A. *Methodology for quantifying the cost effectiveness of freeway service patrols programs—Case study: H.E.L.P. program*. 2005.
17. Chung, Y. Identifying Primary and Secondary Crash from Spatio-temporal Crash Impact Analysis. No. 1148, 2013, pp. 1–15.
18. Yang, H., B. Bartin, and K. Ozbay. Identifying Secondary Crashes on Freeways Using Sensor Data. 2013.

19. Yang, H., K. Ozbay, E. F. Morgul, B. Bartin, and K. Xie. Development of an On-line Scalable Approach for Identifying Secondary Crashes. 2014.
20. Owens, O. *Traffic Incidents on the M1 Motorway in Hertfordshire*. Crowthorne, Berkshire, England, 1978, pp. 1–14.
21. Zhang, H., and A. Khattak. What Is the Role of Multiple Secondary Incidents in Traffic Operations? *Journal of Transportation Engineering*, Vol. 136, No. 11, 2010, pp. 986–997.
22. Vlahogianni, E. I., M. G. Karlaftis, J. C. Golias, and B. M. Halkias. Freeway Operations, Spatiotemporal-Incident Characteristics, and Secondary-Crash Occurrence. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2178, Dec. 2010, pp. 1–9.
23. Khattak, A. J., X. Wang, and H. Zhang. Spatial Analysis and Modeling of Traffic Incidents for Proactive Incident Management and Strategic Planning. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2178, 2010, pp. 128–137.
24. Vlahogianni, E. I., M. G. Karlaftis, and F. P. Orfanou. Modeling the Effects of Weather and Traffic on the Risk of Secondary Incidents. *Journal of Intelligent Transportation Systems*, Vol. 16, No. 3, 2012, pp. 109–117.
25. Jodoin, P. Traffic Incident Management. http://ops.fhwa.dot.gov/eto_tim_pse/about/tim.htm#t3. Accessed Oct. 3, 2014.
26. Hamzehei, A., E. Chung, and M. Miska. Traffic safety risks trends and patterns analysis on motorways. 2014.
27. Zhou, H., R. Fries, M. Williamson, Z. Tian, and B. Vaughn. *Development of A Highway Incident Management Operational and Training Guide*. 2011, pp. 1–49.
28. Zhang, H., and A. J. Khattak. Analysis of Cascading Incident Event Durations on Urban Freeways. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2178, 2010, pp. 30–39.
29. Federal Highway Administration. 2009 Traffic Incident Management (TIM) Self-Assessment (SA) Guide. http://www.ops.fhwa.dot.gov/eto_tim_pse/docs/09timsaguide/index.htm.
30. Carson, J. L. *Best Practices in Traffic Incident Management*. Washington, D.C., 2010, pp. 1–123.
31. Latoski, S. P., R. Pal, and K. C. Sinha. Cost-Effectiveness Evaluation of Hoosier Helper Freeway Service Patrol. *Journal of Transportation Engineering*, Vol. 125, No. 5, 1999, pp. 429–438.
32. Guin, A., C. Porter, B. Smith, and C. Holmes. Benefits Analysis for Incident Management Program Integrated with Intelligent Transportation Systems Operations: Case Study. *Transportation Research Record*, Vol. 2000, No. 1, Jan. 2007, pp. 78–87.
33. Ng, M., A. Khattak, and W. Talley. Modeling the time to the next primary and secondary incident: A semi-Markov stochastic process approach. *Transportation Research Part B: Methodological*, Vol. 58, 2013, pp. 44–57.
34. Freeway Service Patrol.
35. Sohrweide, T. A. FIRST Program Evaluation. 2005.

36. Hagen, L., H. Zhou, and H. Singh. *Road Ranger Benefit Cost Analysis*. Tampa, Florida, 2005, pp. 1–76.
37. Lin, P.-S., A. Fabregas, and H. Chen. Cost Benefit Analysis of Freeway Service Patrol Programs: A Case Study in Florida. 2012.
38. Lin, P.-S., A. Fabregas, H. Chen, H. Zhou, and Q. Wang. *Review and Update of Road Ranger Cost Benefit Analysis*. Tampa, Florida, 2012, pp. 1–89.
39. Dougald, L. E., and M. J. Demetsky. *Performance Analysis of Virginia's Safety Service Patrol Programs: A Case Study Approach*. Charlottesville, Virginia, 2006, pp. 1–76.
40. Yin, Y. Optimal fleet allocation of freeway service patrols. *Networks and Spatial Economics*, Vol. 6, No. 3, 2006, pp. 221–234.
41. Parker, S. T., and Y. Tao. WisTransPortal: A Wisconsin Traffic Operations Data Hub. 2006.
42. Parker, S. WisTransPortal Developer Wiki. http://transportal.cee.wisc.edu/resources/wiki/. Accessed Jan. 7, 2013.
43. Wisconsin Department of Transportation. Official state trunk highway system maps. http://www.dot.wisconsin.gov/travel/maps/sth.htm. Accessed Jan. 7, 2013.
44. Dijkstra, E. W. A Note on Two Problems in Connexion with Graphs. *Numerische Mathematik*, Vol. 1, 1959, pp. 269–271.
45. Zhang, H., and A. Khattak. Spatiotemporal Patterns of Primary and Secondary Incidents on Urban Freeways. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2229, 2011, pp. 19–27.
46. Yang, H., B. Bartin, and K. Ozbay. Investigating the Characteristics of Secondary Crashes on Freeways. 2013.
47. Imprialou, M. M., F. P. Orfanou, E. I. Vlahogianni, and M. G. Karlaftis. Defining Spatiotemporal Influence Areas in Freeways for Secondary Accident Detection. 2013.
48. Hurdle, V., and B. Son. Shock Wave and Cumulative Arrival and Departure Models: Partners Without Conflict. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1776, 2001, pp. 159–166.
49. Li, H., and R. Bertini. Comparison of Algorithms for Systematic Tracking of Patterns of Traffic Congestion on Freeways in Portland, Oregon. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2178, No. 1, 2010, pp. 101–110.
50. Saddi, R. R. *Studying the Impacts of Primary Incidents on Freeways to Identify Secondary Incidents*. University of Nevada, Las Vegas, 2009.