

Secondary Crash Identification on A Large-scale Highway System

Dongxi Zheng (corresponding author)

Research Assistant

TOPS Laboratory, Department of Civil and Environment Engineering

University of Wisconsin-Madison

1415 Engineering Drive, Room 1249A, Madison, WI, 53706

Phone number: (608)335-0889, E-mail: dzheng3@wisc.edu

Madhav V. Chitturi, PhD

Assistant Researcher

TOPS Laboratory, Department of Civil and Environment Engineering

University of Wisconsin-Madison

B243 Engineering Hall, 1415 Engineering Drive, Madison, WI 53706

Phone number: (608) 890-2439, Fax: (608)262-5199, E-mail: mchitturi@wisc.edu

Andrea R. Bill

Associate Researcher

TOPS Laboratory, Department of Civil and Environment Engineering

University of Wisconsin-Madison

B243 Engineering Hall, 1415 Engineering Drive, Madison, WI 53706

Phone number: (608) 890-3524, Fax: (608)262-5199, E-mail: bill@wisc.edu

David A. Noyce, PhD, PE

Associate Professor

Director, Traffic Operations and Safety (TOPS) Laboratory

University of Wisconsin-Madison

Department of Civil and Environment Engineering

1204 Engineering Hall, 1415 Engineering Drive

Madison, WI 53706

Phone number: (608) 265-1882, Fax: (608)262-5199, E-mail: noyce@engr.wisc.edu

Submission date: 11/14/2013

Word count: Text (6371) + 250*Figure (4) + 250*Table (0) = 7371

1 **ABSTRACT**

2 The annual cost of congestion in the United States reportedly exceeds \$120 billion (1).
3 Freeway incidents are major sources of non-recurrent congestion and the resulting secondary
4 crashes can prolong the traffic impact and increase the cost. Research on secondary crashes to
5 support statewide transportation system management has been limited. In the current study, a
6 two-phase automated procedure is developed to identify secondary crashes on large scale
7 regional transportation systems. In the first phase, a crash pairing algorithm is developed to
8 extract spatially and temporally near-by crash pairs. The accuracy and efficiency of the algorithm
9 were validated by comparing to an ArcGIS based program. In the second phase, two filters are
10 proposed to reduce the crash pairs for secondary crash identification: the first filter selects crash
11 pairs whose earlier crashes were on mainline highways; the second filter selects crash pairs
12 whose later crashes happened within the dynamic impact areas (i.e., backup queues) of the
13 earlier crashes. Shockwave theory is used to model the dynamic impact of a primary incident.
14 The two-phase procedure uses a linear referencing system for crash localization and can be
15 applied to any regional transportation system with a similar data structure. A case study was
16 conducted on nearly 1,500 miles of freeways in Wisconsin using 2010 data. Among the crash
17 pairs produced by the two-phase procedure, 79 secondary crashes were confirmed via police
18 reports. Preliminary analyses showed that 1) secondary crashes occurred in the same traffic
19 direction as the primary incidents were about three times greater in frequency compared to
20 secondary crashes in the opposing direction, and 2) two-vehicle rear-ends, multiple-vehicle rear-
21 ends, and sideswipes were three major types of secondary crashes (over 85%).

1 INTRODUCTION

2 A secondary crash is an undesirable consequence resulting from a primary incident. More
3 formally, according to the Federal Highway Administration (FHWA), “secondary crashes are
4 those that occur with the time of detection of the primary incident where a collision occurs either
5 a) within the incident scene or b) within the queue, including the opposite direction, resulting
6 from the original incident” (2). Existing studies have shown the extended traffic impact and the
7 economic costs of secondary crashes (3–5). Reducing the chances of secondary crashes becomes
8 a major consideration in the dispatch plans of traffic incident management (TIM) agencies (6, 7).

9 In spite of various findings on secondary crashes, most existing studies were limited by
10 scope. Many studies were conducted on only one or two sample freeways or a short segment of
11 highway; other studies extended the scope to freeways but considered a small regional scale.
12 Only two studies were performed at a large scale that involved statewide highway systems. One
13 of the major reasons for such scope constraints was the challenge of secondary crash
14 identification. In order to identify secondary crashes accurately, most existing studies considered
15 the dynamic features of the traffic impact caused by the primary incidents. Thus, the study
16 scopes were limited to highway facilities where high resolution traffic data were available for
17 dynamic analyses. In addition, modeling the dynamic impact of primary incidents required
18 considerable computational efforts, which for a statewide transportation system could be
19 intolerable or even infeasible. Previous studies considering statewide highway systems did not
20 consider the dynamic impact of primary incidents. In summary, none of the previous studies
21 investigated secondary crashes on a statewide transportation system while considering the
22 dynamic impact of primary incidents.

23 To fill the above research gap, the current study develops a two-phase automatic
24 procedure. In the first phase, spatially and temporally near-by crash pairs (up to custom static
25 thresholds) are extracted from a large network based on a crash pairing algorithm. The accuracy
26 and the efficiency of this algorithm were validated. In the second phase, two filters are used to
27 select crash pairs that are more likely to be primary-secondary crash pairs. One of the filters
28 utilizes shockwave theory to evaluate the dynamic traffic impact of the primary incidents. At the
29 end of the two-phase procedure, manual review of identified police reports is needed to confirm
30 actual secondary crashes. However, the number of crash reports to review is considerably less.

31 LITERATURE REVIEW

32 Secondary crashes have been observed to be one of the notable consequences of freeway
33 incidents. Early in 1970s, Owens conducted an on-the-spot study of traffic incidents on a 21
34 kilometer (13 mile) stretch of motorway in England during peak hours and found that 32.5% of
35 the observed crashes were related to primary incidents (8). In recent decades, the development of
36 intelligent transportation systems (ITS) has made a variety of transportation data easier to access,
37 which in turn has encouraged researchers to revisit secondary crashes. In earlier studies (3–5, 9–
38 17), an incident was identified as a secondary crash as long as it occurred within a rectangle
39 time-space window originated from another incident. For example, Raub classified an incident as
40 a secondary crash if it happened within 1,600 meters upstream of another incident and no later
41 than 15 minutes after that incident was cleared (9, 10). This method type was called static
42 thresholds in the sense that it considered the spatial impact range of a primary incident to be
43 consistent throughout a certain time period. However, the impact of a traffic incident is typically
44 dynamic with respect to time. Later studies incorporated this fact in secondary crash

1 identification (18–29). The earliest attempt was made by Moore et al. who classified an incident
2 as a secondary crash only if it fell under the progression curve (i.e., the resulting queue boundary
3 as a function of time based on real-time queue end tracking data) of another incident (5). Curves
4 with a similar concept were generated using the traffic arrival-departure model in other studies
5 (20–22, 24). In fact, these dynamic curves only depicted the moving queue fronts, but did not
6 consider the queue release from the incident location since the onset of incident clearance. To
7 accommodate the releasing front, researchers have used either the speed contour map method or
8 the ASDA model to depict the impact area of an incident (19, 23, 25–29). However, shockwave
9 theory, which can also model the queuing and the releasing dynamics, has not been used in the
10 literature for secondary crash identification.

11 Research on secondary crashes at the scale of a statewide transportation system has been
12 limited. A majority of the existing studies focused on one or two sample freeways or only a
13 stretch of a highway of which detailed traffic conditions could be obtained through densely
14 deployed traffic detectors, closed circuit traffic cameras (CCTV), or even aircraft based
15 congestion tracking systems (4, 14, 16, 18, 19, 23, 25–28). Some other studies extended the
16 research scope to several freeways or urban arterials within a fully patrolled and ITS assisted
17 district (3, 5, 9, 10, 12, 13, 20–22, 24, 29). Only a few studies were conducted on statewide
18 transportation systems (11, 17). Identifying all spatially and temporally near-by crash pairs from
19 a large highway network, and hence a significant amount of input crashes, could be
20 computationally complex. None of the above studies provided an efficient procedure.

21 Based on the above literature review, two primary objectives of the current study were set.
22 First, an efficient algorithm to identify all near-by crash pairs (up to custom static thresholds) for
23 a statewide transportation system should be developed. Second, in order to reduce the candidate
24 primary-secondary crash pairs based on the static thresholds, additional filters that incorporate
25 the dynamic feature of primary incident impact should be established.

26 DATA DESCRIPTION

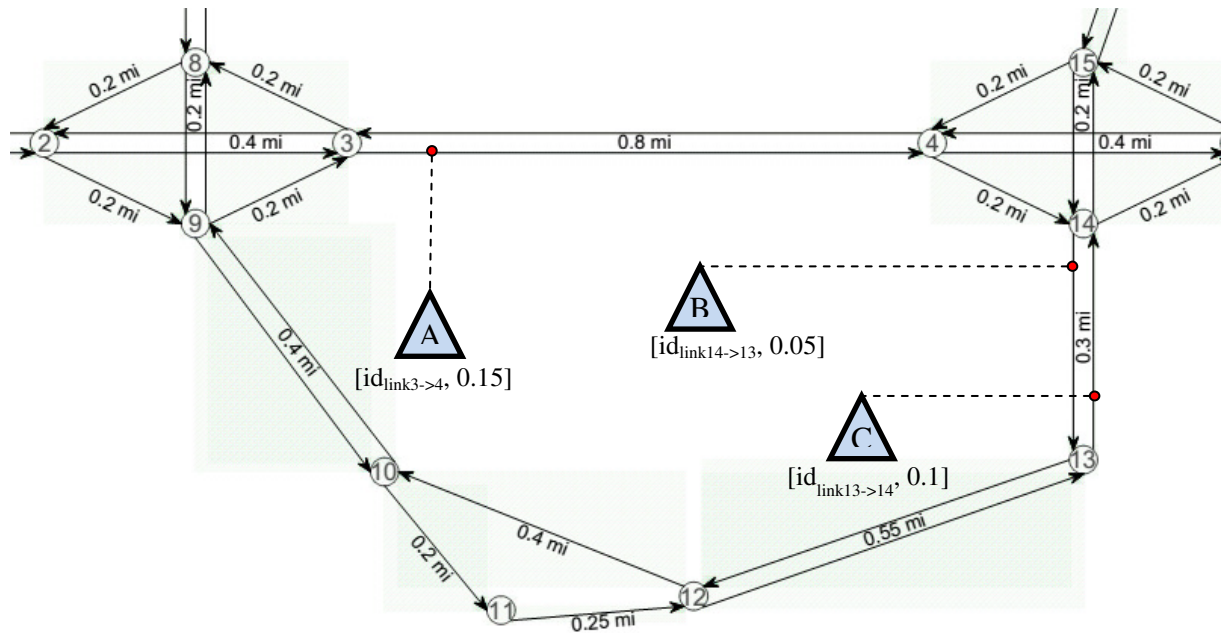
27 The WisTransPortal Data Hub of the Traffic Operations and Safety (TOPS) Laboratory at
28 the University of Wisconsin – Madison houses a variety of statewide transportation data prepared
29 and provided by the Wisconsin Department of Transportation (WisDOT) (30, 31). Among these
30 data, the State Trunk Network (STN) data and the crash data are the two primary inputs to the
31 current study. The STN includes the state trunk highways (STH), the US highways (US), the
32 interstates highways (IH), the designated freeways, and the expressways in Wisconsin as of 2012
33 (32). The crash data cover all reported crashes in Wisconsin since 1994 and are updated monthly.
34 WisDOT provides both the maps (i.e., Esri shapefiles) and the database records of STN and
35 crashes. WisDOT also embeds a linear referencing system in the crash records to allow locating a
36 crash on the STN without using the maps. For the proposed algorithm, the database records with
37 the linear referencing system were used. The maps were used for validation and comparison
38 purposes.

39 STN Links and Linear Referencing

40 A traditional way of modeling highway networks is a figure that consists of nodes and
41 directional links. The STN is stored in this manner. Nodes in STN are called reference sites (RS).
42 Each link in the STN starts from one RS (RS^{from}) and ends at another RS (RS^{to}). A link represents
43 a highway segment, either mainline or ramp, in one traffic direction with relatively consistent

1 geometric layout (e.g., number of lanes, lane width, etc.). Figure 1 illustrates how a small portion
 2 of a highway network is represented in the STN. Links are displayed as solid arrows with their
 3 lengths. RS's are labeled in circles. An arbitrary location on the STN, for example a crash scene,
 4 is determined by a linear coordinate $[link\ id, link\ offset]$, namely linear referencing. The $link\ id$
 5 tells in which link the crash occurred and the $link\ offset$ tells the distance from the link's RS^{from}
 6 to the crash. As of the year 2012, the total number of in-operation links is 33,015 and the total
 7 length is 24,903 miles.

8



Example distances:

$$d(A,B) = (0.8 - 0.15) + 0.2 + (0.3 - 0.05) = 1.1 \text{ miles}$$

$$d(B,C) = 0.3 - 0.05 - 0.1 = 0.15 \text{ miles}$$

9

10 **FIGURE 1 Illustration of the linear referencing system.**

11 **Crash Records**

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

18 **Other Data**

19 In addition to the STN links, WisTransPortal also stores other highway information. For
 20 example, all the routes in STN are stored in a table, each record representing the entire stretch of
 21 a highway route and its geographical direction (e.g., US 12 East Bound); virtual mile markers are
 22 stored as reference points. Traffic data are also available. WisDOT manages the TRAFFIC DATA

1 System (TRADAS) and the Advanced Transportation Management System (ATMS), with traffic
 2 detectors deployed on the STN. WisTransPortal contains information of these TRADAS and
 3 ATMS detectors as well as their traffic counts.

4 **FIRST PHASE: CRASH PAIRING ALGORITHM**

5 Given the STN linear referencing system and the crash records, the target of the crash
 6 pairing algorithm is to identify all crash pairs (c_i, c_j) that satisfy Formulas 1 and 2. In Formula 2,
 7 $d(c_i, c_j)$ is measured along the STN links by treating the links as bi-directional (see examples in
 8 Figure 1). Highway splits, merges, and intersections should be accommodated, which was not
 9 addressed by previous studies focusing on individual freeways.

10

$$0 \leq t(c_j) - t(c_i) \leq T \quad (1)$$

$$d(c_i, c_j) \leq D \quad (2)$$

where,

c_i = Crash i , the former crash;

c_j = Crash j , the later 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.

11

12 Given the significant sizes of the STN links and the crashes, simple algorithms are either
 13 slow or infeasible. One naïve algorithm is to run Dijkstra's method repeatedly for every crash.
 14 Dijkstra's method is an iterative approach that finds the shortest path from an origin to every
 15 node in a network. Dijkstra's method can be briefly summarized as follows: All nodes are
 16 considered to be infinitely distant from the origin and "unvisited" initially. The method begins
 17 from the origin and computes the distances to its neighbors (i.e., nodes with direct connection)
 18 and marks the origin as "visited". In every successive iteration step, the method chooses the
 19 closest "unvisited" node to the origin, updates the distances from the origin to that node's
 20 "unvisited" neighbors if the paths become shorter through that node, and marks that node as
 21 "visited". The iteration continues until every node is "visited". At the end, the distances from the
 22 origin to each node are the shortest distances (33). The complexity of Dijkstra's method is $O(N^2)$
 23 in respect to N crashes, where N is larger than 100,000 for an annual statewide analysis. By
 24 repeatedly using Dijkstra's method for N crashes, the complexity of the naïve algorithm becomes
 25 $O(N^3)$, which is not efficient. Another alternative is to use dynamic programming to populate a
 26 shortest path matrix between every two crashes. This alternative is infeasible because it not only
 27 spends an equivalent amount of computation time as the first algorithm but also requires
 28 unacceptable computer memory space (e.g., $100,000^2 * 8 \text{ bytes} \approx 75 \text{ GB}$) to store the matrix.

29 The proposed pairing algorithm first analyzes the relationships between links and uses
 30 these relationships to derive crash-to-crash distances. For each link, lk_i , that contains one or more
 31 crashes, the algorithm performs a variant of Dijkstra's traversal (as will be explained later) and
 32 generates the relationships between lk_i and the other links. The distances between crashes are

1 then calculated based on these relationships. Compared to the first algorithm mentioned above,
 2 the number of traversals is bounded by the total number of links no matter how many crashes are
 3 analyzed. The pairing algorithm also utilizes the D mile space window to constrain the Dijkstra's
 4 traversal to a relevant portion (normally small) of the STN network. In the following subsections,
 5 the concept of a local linear coordinate system is introduced, based on which the relationship
 6 between two links can be comprehensively defined. Additionally, the equation to derive crash-
 7 to-crash distance from the link-to-link relationship is also given, along with the concept of a
 8 candidate link that is used to constrain the Dijkstra's traversals, the pseudo code of the algorithm
 9 with special case explanation, and finally, the validation of this algorithm.

10 Local Linear Coordinate System

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

- 15 • **Forward (positive) coordinate** (x_{RS}^+) = the length of the *base link* + $d(RS_{base}^{to},$
 16 $RS) | RS_{base}^{from}$. $d(RS_{base}^{to}, RS) | RS_{base}^{from}$ is the shortest network distance between RS_{base}^{to}
 17 and RS in a sub-network without RS_{base}^{from} (and links connected to it). If $d(RS_{base}^{to},$
 18 $RS) | RS_{base}^{from}$ does not exist, $x_{RS}^+ = +\infty$. Specially, $x_{RS_{base}^{from}}^+$ is defined as 0.
- 19 • **Backward (negative) coordinate** (x_{RS}^-) = $d(RS_{base}^{from}, RS) | RS_{base}^{to}$. $d(RS_{base}^{from},$
 20 $RS) | RS_{base}^{to}$ is the shortest network distance between RS_{base}^{from} and RS in a sub-
 21 network without RS_{base}^{to} (and links connected to it). If $d(RS_{base}^{from}, RS) | RS_{base}^{to}$ does
 22 not exist, $x_{RS}^- = +\infty$. For example, $x_{RS_{base}^{to}}^- = +\infty$.

23 As an example, in Figure 1, consider RS_{12} under the LLCS of $link_{3 \rightarrow 4}$ (as the base link).
 24 $x_{RS_{12}}^+ = 0.8 (link_{3 \rightarrow 4}) + 0.2 (link_{4 \rightarrow 14}) + 0.3 (link_{14 \rightarrow 13}) + 0.55 (link_{13 \rightarrow 12}) = 1.85$ miles. $x_{RS_{12}}^- = 0.2$
 25 $(link_{9 \rightarrow 3}) + 0.4 (link_{10 \rightarrow 9}) + 0.4 (link_{12 \rightarrow 10}) = 1.0$ mile.

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

32 In the context of a LLCS, any link (including the base link) is related to the base link by
 33 the LLCS coordinates of its RS^{from} and RS^{to} . Specifically, let a link to be related to the base link
 34 be called a **test link** and its end RS's be denoted as RS_{test}^{from} and RS_{test}^{to} . Vector $v_{test} =$
 35 $\left[x_{RS_{test}^{from}}^+, x_{RS_{test}^{from}}^-, x_{RS_{test}^{to}}^+, x_{RS_{test}^{to}}^- \right]$ is defined as the relationship vector of the test link in the LLCS.
 36 With the relationship vector, the network distance between a crash c_{base} on the base link and a
 37 crash c_{test} on a test link can be easily calculated using Equations 3-7. Since the four coordinates
 38 in the relationship vector might result from different routings, there could be four possible crash-

1 to-crash distances (Equations 4-7) whose geometric meanings are demonstrated in Figure 2. The
 2 final crash-to-crash distance should be the smallest possible distance. Besides the distance value,
 3 one can also tell if the two crashes were in the same traffic direction. For example, if the final
 4 distance is d_F^+ (upper right case in Figure 2), the centerline of the resulting route is bolded and
 5 the traffic directions of both crashes (green arrows) are on the same side of the centerline,
 6 meaning the two crashes (or links) were in the same traffic direction; otherwise, like d_F^- and d_T^+ ,
 7 the two crashes were in the opposite traffic directions. Additionally, one can also determine
 8 whether c_{test} happened upstream or downstream of c_{base} . For instance, c_{test} happened upstream of
 9 c_{base} if d_T^+ or d_T^- is the final distance (when the test crash direction follows the bolded route);
 10 otherwise, c_{test} happened downstream of c_{base} (when the test crash direction departs the bolded
 11 route).

$$d(c_{base}, c_{test}) = \min(d_F^+, d_T^+, d_F^-, d_T^-) \quad (3)$$

$$d_F^+ = x_{RS_{test}^{from}}^+ - os_{base} + os_{test} \quad (4)$$

$$d_T^+ = x_{RS_{test}^{to}}^+ - os_{base} + (l_{test} - os_{test}) \quad (5)$$

$$d_F^- = x_{RS_{test}^{from}}^- + os_{base} + os_{test} \quad (6)$$

$$d_T^- = x_{RS_{test}^{to}}^- + os_{base} + (l_{test} - os_{test}) \quad (7)$$

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;

$x_{RS_{test}^{from}}^-$ = Backward coordinate of RS_{test}^{from} , mile;

$x_{RS_{test}^{to}}^+$ = Forward coordinate of RS_{test}^{to} , mile;

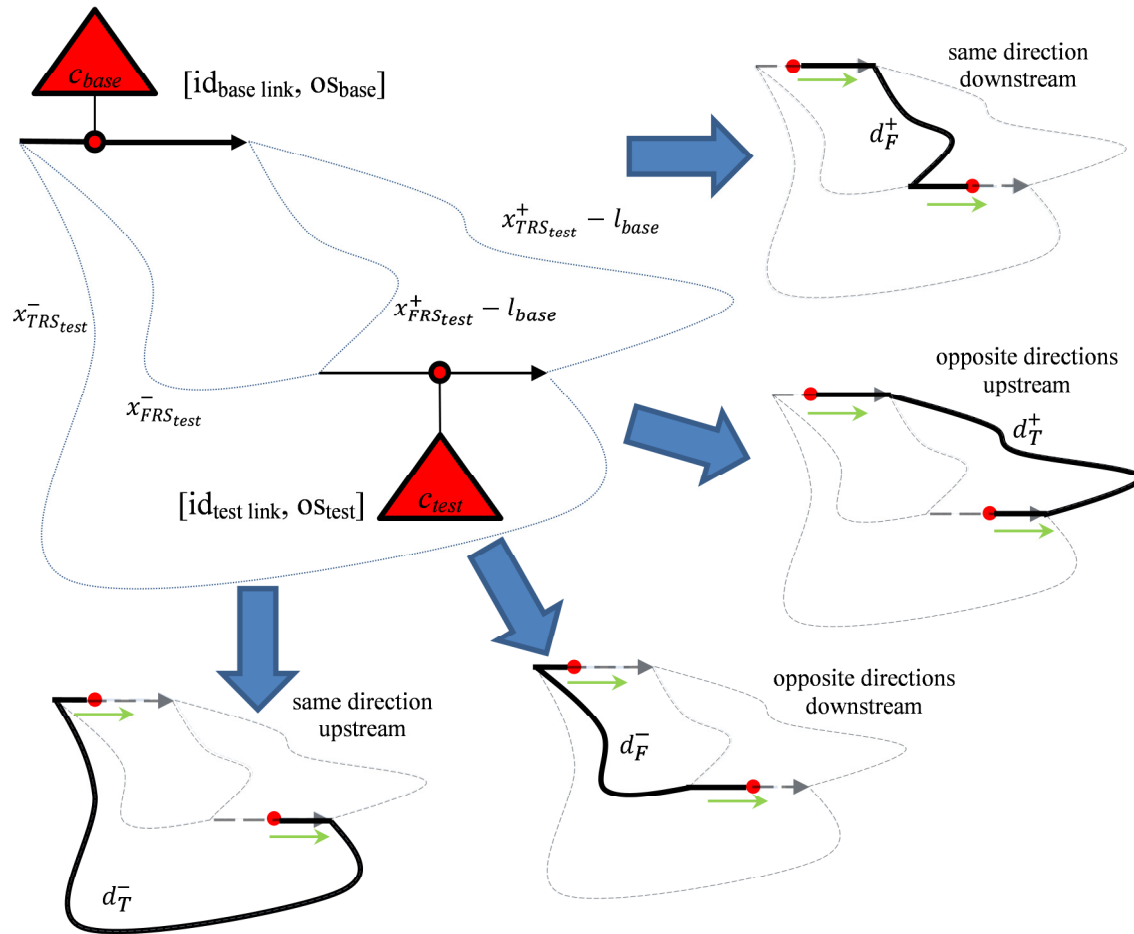
$x_{RS_{test}^{to}}^-$ = Backward coordinate of RS_{test}^{to} , mile;

os_{base} = Link offset of c_{base} , mile;

os_{test} = Link offset of c_{test} , mile;

l_{test} = Length of the test link, mile.

12



1
2 **FIGURE 2 Four possible distances between two crashes based on the relationship vector.**

3 **Candidate Link**

4 In the previous section, every link is assumed to be tested against the base link. However,
5 given a particular spatial threshold of D miles, a test link too far away from the base link is
6 irrelevant to finding the near-by crash pairs. Only those links whose relationship vectors satisfy a
7 certain condition may contain crashes within D miles of the base link crashes. In fact, the
8 condition is as simple as $\min(x_{RS_{test}^{from}}^+ - l_{base}, x_{RS_{test}^{from}}^-, x_{RS_{test}^{to}}^+ - l_{base}, x_{RS_{test}^{to}}^-) \leq D$, where l_{base}
9 is the length of the base link. Links satisfying this condition are called candidate links and form a
10 relatively small and relevant portion of the network (when D is relatively small). The two passes
11 of Dijkstra’s traversal can stop expansion as early as any further RS to be reached has a forward
12 coordinate larger than $l_{base} + D$ and a backward coordinate larger than D . Then, all links
13 connected to the already reached RS’s are all the candidate links.

14 **The Algorithm**

15 Below is the pseudo code of the proposed crash pairing algorithm. l_{base} is assumed to be
16 a preprocessed set of links containing at least one crash. The statement “find all candidate links”
17 refers to the preparation of the relationship vectors for all candidate links in the LLCS as
18 described in the above sections. $t(*)$ is the function of getting the time of a crash in minutes since

1 a consistent time origin. T and D are the static thresholds in minutes and miles, respectively. It
 2 should be noted that the recorded time of crash could be slightly different from the time when the
 3 crash actually occurred. However, the authors do not expect it to have a significant impact on the
 4 results since a large time threshold of 5 hours was used. The statement “calculate $d(c_{base}, c_{cand})$ ”
 5 refers to Equations 3-7.

6 For each lk_{base} in L_{base} :
 7 Find all candidate links of lk_{base} as a set L_{cand} ;
 8 For each candidate link lk_{cand} in L_{cand} :
 9 For each crash c_{base} in lk_{base} :
 10 For each crash c_{cand} in lk_{cand} :
 11 If $0 \leq t(c_{cand}) - t(c_{base}) \leq T$:
 12 Calculate $d(c_{base}, c_{cand})$;
 13 If $d(c_{base}, c_{cand}) \leq D$:
 14 Add (c_{base}, c_{cand}) as a pair in the result;

15
 16 A special case should be treated differently. As illustrated in Figure 1, the longitudinal
 17 distance between crash B and crash C was only 0.15 miles while they occurred on opposite sides
 18 of the same highway. However, relying only on the network traversal of links, the resulting
 19 distance will go around RS₁₄ and be calculated as 0.25 miles. Such unrealistic result is not
 20 desirable. In order to overcome this limitation, additional information from the STN were
 21 employed. An STN table of route-links is used to aid the links with their physical meanings.
 22 Each record in the route-link table tells which highway a link belongs to in what direction. All
 23 links on the other side of the same highway are considered candidate links of the base link. When
 24 calculating the distance between a crash on the base link and a crash on the other side of the
 25 highway, the algorithm calculates the cumulative distances from the two crashes to a far
 26 upstream/downstream shared RS on the highway. The difference between these two cumulative
 27 distances is considered the distance between these two crashes. Additionally, when a shared RS
 28 could not be found, the algorithm further utilize another set of highway reference locations,
 29 reference points (RP). Each RP has its on-highway number, on-highway direction, RP number,
 30 and RP letter. If two RP's have the same on-highway number, RP number, and RP letter, they
 31 correspond to the same longitudinal position on the highway, even with different on-highway
 32 directions. Additionally, each RP, like a crash location, has a linear reference that maps it on to a
 33 link. Based on the above input, if two links on opposite sides of the same highway contains RP's
 34 with the same RP number and RP letter, there is a shared longitudinal position between them.
 35 Thus, instead of looking for a shared RS, the algorithm looks for a shared longitudinal position
 36 based on RP's.

37 **Validation**

38 The pairing algorithm was implemented as a Java program. The program passed several small
 39 independent tests (e.g., the entire stretch of a particular highway in STN with crashes of several
 40 days) with manual extracted ground truths. In order to further validate the accuracy and the

1 efficiency of the algorithm, a large scale network was tested. Since the ground truth in the large
2 scale test was infeasible to be manually extracted, a relatively reliable ArcGIS based program
3 was used as a mutual validation reference. The basic idea of the ArcGIS based program is to
4 prepare a network dataset using the STN shapefile and the crash shapefile and use the buffer
5 function of the NetworkAnalyst toolbox to find, for each crash, every other crash that is within a
6 buffer network distance (the spatial threshold) from that crash. The ArcGIS based program was
7 implemented in C++ using the ArcGIS APIs. Due to the unavailability of control over the buffer
8 function of the NetworkAnalyst toolbox, the ArcGIS based program was similar to the naïve
9 algorithm of traversing the network for every pair of crashes, which provided the authors a
10 chance to compare the efficiencies.

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

22 **SECOND PHASE: CRASH PAIR FILTERS**

23 For the purpose of identifying secondary crashes, the pairing algorithm produces an
24 initial searching set, which without additional filtering could be too vast to be useful. Two filters
25 are proposed below as to select out crash pairs that are more likely to capture primary-secondary
26 relationship.

27 **Ramp Filter**

28 A crash pair resulting from the proposed algorithm will be excluded if its former crash
29 happened on a highway on-ramp or a highway off-ramp. A crash is determined to be on a ramp if
30 the link on which it happened represents a portion or entire segment of a ramp. The rationale
31 behind this filter is that ramp crashes rarely caused secondary crashes. To evaluate this
32 assumption, 85 sample crash pairs whose former crashes happened on ramp were selected and
33 verified. One or two crash pairs were randomly sampled from each 1-hour-5-mile intervals of the
34 5-hour-10-mile thresholds. Manual review showed that none of the 85 samples contained a
35 primary-secondary pair. Although one crash pair involved two secondary crashes, but they were
36 not related to each other; in addition, these two secondary crashes were captured by the actual
37 primary-second pairs whose primary crashes were not on ramps.

38 **Impact Area Filter**

39 As mentioned in the literature review, crash pairs resulted from static thresholds could
40 contain false primary-secondary pairs. These false pairs generally have unreasonable
41 combinations of time distance and spatial distance. For example, a candidate pair whose time
42 distance is 0 minutes but the spatial distance is 5 miles is certainly not a primary-secondary crash
43 pair. Since secondary crashes have been recognized to be in the queue caused by the primary

1 incidents, queue theories were commonly used to establish the time-varying impact area of the
 2 primary incidents to identify secondary crashes (13, 14, 18–29). Comparison of various queue
 3 estimation methods can be found in more general traffic research papers (34, 35). Based on the
 4 literature review, none of the previous secondary crash studies used the shockwave model to
 5 estimate the impact area (IA) caused by a primary incident. In the current study, the IA of a crash
 6 is defined between two simplified straight shockwave lines, one for the queuing shockwave and
 7 the other for the discharging shockwave (Figure 3). Mathematical representation to judge if a
 8 crash fell into the IA of another is given in Equations 8 through 10. Traffic flow of the prevailing
 9 traffic condition (q_i) is the monthly average hourly traffic volume provided by the TRADAS
 10 detectors, with the same day of week and the same hour of day as the former crash. If the later
 11 crash happened outside the IA and its parallel portion on the opposite traffic direction of the
 12 former crash, the crash pair will be excluded. On the other hand, secondary crashes could happen
 13 in the vicinity of the primary incident during its clearance. This type of secondary crashes was
 14 typically attributed to the “rubbernecking” effect (8, 36). In order to capture these secondary
 15 crashes, a crash pair whose spatial distance (upstream or downstream in either traffic direction)
 16 was no larger than 1 mile and whose temporal distance was no large than 1 hour should be
 17 reserved even if it does not satisfy the IA requirement.

$$a_2 \times (t - t_{clearance}) \leq d \leq a_1 \times t \quad (8)$$

$$a_1 = (q_2 - q_1)/(k_2 - k_1) \quad (9)$$

$$a_2 = (q_3 - q_2)/(k_3 - k_2) \quad (10)$$

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).

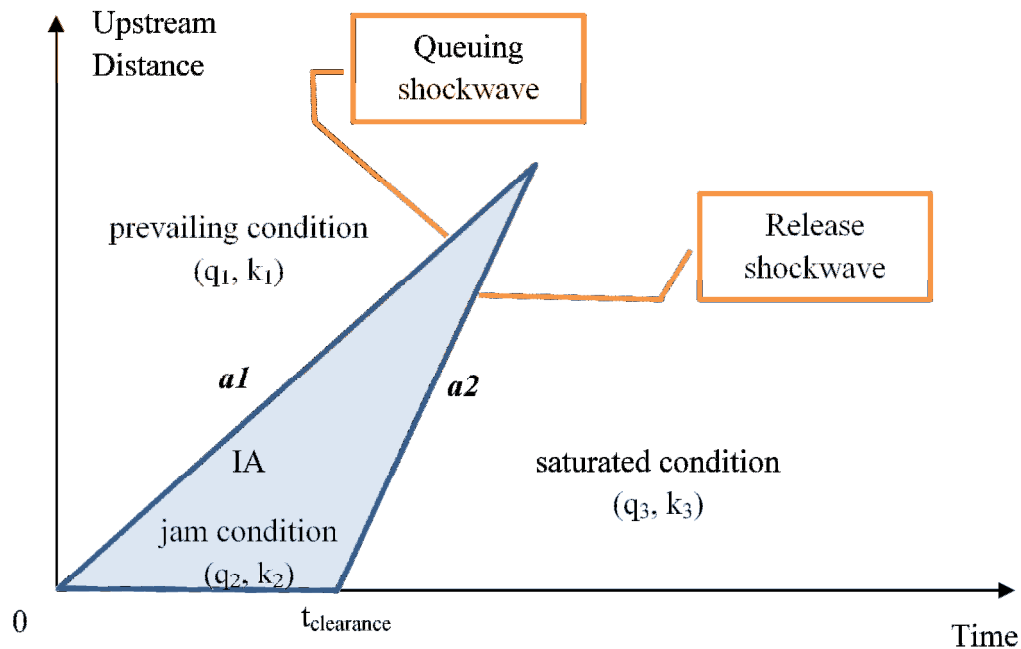


FIGURE 3 Illustration of an impact area.

CASE STUDY

A case study was performed on crashes happened on approximately 1,500 miles of freeways in Wisconsin for the year 2010. The layout of these freeways in relation to the entire STN network was illustrated in the upper-right area of Figure 4. Although the case study only used crashes happening on these freeways, the calculation of network distances was not subjected to these freeways, but instead, relied on the entire STN network. A total of 12,513 raw input crashes were retrieved for year 2010, the last five hours of 2009, and the first five hours of 2011. The inclusion of five hours into the previous year and the next year corresponds to the selected 5 hour temporal threshold so crash pairs crossing the new year's boundary could be captured. The workflow and the resulting reduced data in each step are summarized in the left area of Figure 4.

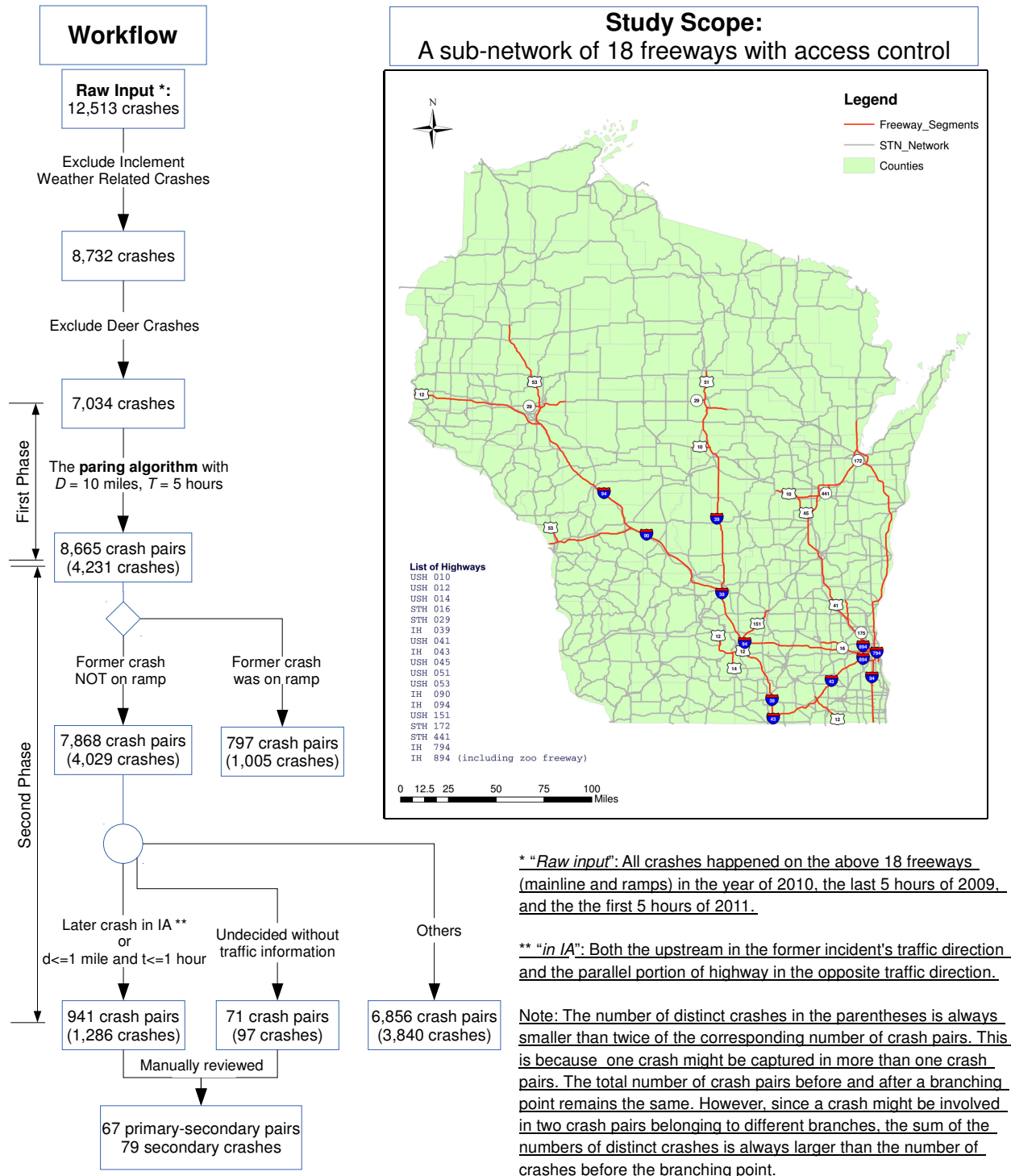
Before applying the pairing algorithm, the raw input crashes were first reduced based on a focused study scope that excluded inclement weather conditions and deer crashes. In Wisconsin, a large portion of crashes were related to inclement weather during the winter. For example, in January 2010, 1,520 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 to contribute to 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. For a similar reason, deer crashes (about 20% of all crashes) were excluded from the study. As a result, 7,034 crashes remained as the input to the proposed algorithm.

Conservative thresholds of 10-mile-5-hour were used for the first phase (crash pairing).

1 The 10-mile-5-hour thresholds are approximately 5 times larger (in each dimension) than most
2 static thresholds used in the literature and supersede all actual temporal-spatial ranges of
3 primary-second pairs found by existing studies (3–5, 9–14, 16–29). Thus, using even larger
4 thresholds is unlikely to include more actual primary-secondary crash pairs. The pairing
5 algorithm generated 8,665 crash pairs (4,231 distinct crashes). The second phase (crash pair
6 filtering) further reduced the number of crash pairs down to 1,012 (88.3% reduction). Up to this
7 point in the analysis, all computations were completed automatically within 2 hours. The
8 resulting 1,012 crash pairs for manual review only contained 1,347 distinct crashes. Compared to
9 the initial input of 7,034 crashes, the review effort was saved by 81%.

10 Secondary crashes and their corresponding primary incidents were confirmed through
11 manual review of police reports. An estimate of 30 man-hours was used in reviewing the 1012
12 candidate crash pairs, averaged to nearly 2 man-minutes per crash pair. Potential employment of
13 ORC (optical character recognition) and artificial intelligence may help to further minimize the
14 reviewing time in the future. A crash was classified as a secondary crash only when its report
15 explicitly referred to a previous crash. This criterion might have resulted in fewer than the actual
16 secondary crashes, but ensured the confidence of further analyses on the resulting secondary
17 crashes. Primary incidents were identified only if they could be matched, by either document
18 number or other key descriptions, to those referred by the secondary crashes. According to these
19 criteria, a total of 79 crash pairs were found to contain secondary crashes. The number of distinct
20 secondary crashes was also 79. Among the 79 pairs, 67 captured the primary incidents.

21 Preliminary analyses were conducted on the resulting primary-secondary pairs and
22 secondary crashes. Among the 67 primary-secondary pairs, 52 secondary crashes (77.6%)
23 happened in the same traffic direction as the primary crashes and the average spatial and
24 temporal distances were 1511 feet and 15.7 minutes, respectively; 15 secondary crashes (22.4%)
25 happened in the opposite traffic direction of the primary crashes and the average spatial and
26 temporal distances were 1264 feet and 18.2 minutes, respectively. Among the 79 secondary
27 crashes, 44 were two-vehicle rear-ends (55.7%), 12 were multiple-vehicle rear-ends (15.2%), 13
28 were sideswipes (16.5%), 5 were hitting debris (7.3%), 2 were angles (2.5%), 2 involved squad
29 vehicles on primary crash scenes (2.5%), and 1 was losing control (1.3%).



1
2

FIGURE 4 Summary of the case study of year 2010.

1 **CONCLUSION, RECOMMENDATION, AND FUTURE WORKS**

2 Secondary crashes are known to prolong the non-recurrent congestion caused by the
3 primary freeway incidents. The benefit of reducing secondary crashes has also been found to
4 exceed the TIM countermeasures such as freeway patrol services. However, research of
5 secondary crashes on large regional transportation systems was limited. The current study
6 contributes to the research community with the following efforts and findings:

- 7 • An efficient crash pairing algorithm was developed to extract spatially and temporally
8 near-by (up to custom static thresholds) crash pairs from a large scale regional
9 transportation system. The accuracy and efficiency of this algorithm were validated.
- 10 • Two effective filters were proposed to select crash pairs that were more likely to capture
11 primary-secondary relationships. The first filter restricts the primary incidents on
12 mainline highways. The second filter restricts the secondary crashes to be within the
13 dynamic impact areas of the primary incidents. Shockwave theory is first used by the
14 current study to estimate the dynamic impact area of a primary incident.
- 15 • A two-phase procedure consisting of the above pairing algorithm and filters automatically
16 narrows down the searching space for secondary crashes in a large regional transportation
17 system. While the procedure is based on the commonly used linear referencing system for
18 crash localization, any transportation system with similar data representation can be
19 analyzed with the procedure. A manual review of the effectively narrowed search space
20 is required.
- 21 • A case study for crashes occurring in 2010 on about 1,500 miles of Wisconsin freeways
22 was conducted. From the crash pairs extracted using the two-phase procedure, 79
23 secondary crashes were confirmed via careful manual review of police reports. Secondary
24 crashes happened in the same traffic direction of the primary incidents were about three
25 times of those occurred in the opposite direction. Two-vehicle rear-ends, multiple-vehicle
26 rear-ends, and sideswipes were three major types of secondary crashes (over 85%). Other
27 crash types, such as hitting debris, angle, losing control, and striking squad vehicles were
28 also observed.

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

38 **ACKNOWLEDGEMENTS**

39 The authors gratefully acknowledge support of this study from Paul Keltner of the
40 Bureau of Traffic Operations, Wisconsin Department of Transportation. Special thanks go to the
41 TOPS lab staff, especially Steven Parker, who manages the WisTransPortal and provided
42 technical support.

1 REFERENCES

- 2 1. Schrank, D., B. Eisele, and T. Lomax. *TTI's 2012 Urban Mobility Report*. 2012.
- 3 2. Margiotta, R., R. Dowling, and J. Paracha. *Analysis, Modeling, and Simulation for Traffic*
4 *Incident Management Applications*. Publication FHWA-HOP-12-045. FHWA, U.S.
5 Department of Transportation, 2012.
- 6 3. Khattak, A. J., X. Wang, and H. Zhang. Are Incident Durations and Secondary Incidents
7 Interdependent? In *Transportation Research Record: Journal of the Transportation Research*
8 *Board*, No. 2099, Transportation Research Board of the National Academies, Washington,
9 D.C., 2009, pp. 39–49.
- 10 4. Karlaftis, M. G., S. P. Latoski, N. J. Richards, and K. C. Sinha. ITS Impacts on Safety and
11 Traffic Management: An Investigation of Secondary Crash Causes. *Journal of Intelligent*
12 *Transportation Systems*, Vol. 5, No. 1, 1999, pp. 39–52.
- 13 5. Moore, James E., II, G. Giuliano, and S. Cho. Secondary Accident Rates on Los Angeles
14 Freeways. *Journal of Transportation Engineering*, Vol. 130, No. 3, 2004, pp. 280–285.
- 15 6. Owens, N., A. Armstrong, P. Sullivan, C. Mitchell, D. Newton, R. Brewster, and T. Trego.
16 *Traffic Incident Management Handbook*. Publication FHWA-HOP-10-013. FHWA, U.S.
17 Department of Transportation, 2010.
- 18 7. Dunn, W. M., and S. P. Latoski. *NCHRP Synthesis of Highway Practice 318: Safe and Quick*
19 *Clearance of Traffic Incidents*. Transportation Research Board of National Academies,
20 Washington, D.C., 2003.
- 21 8. Owens, O. *Traffic Incidents on the M1 Motorway in Hertfordshire*. Transport and Road
22 Research Laboratory, Crowthorne, Berkshire, England, 1978.
- 23 9. Raub, R. A. Secondary Crashes: An Important Component of Roadway Incident
24 Management. *Transportation Quarterly*, Vol. 51, No. 3, 1997, pp. 93–104.
- 25 10. Raub, R. Occurrence of Secondary Crashes on Urban Arterial Roadways. In *Transportation*
26 *Research Record: Journal of the Transportation Research Board*, No. 1581, Transportation
27 Research Board of the National Academies, Washington, D.C., 1997, pp. 53–58.
- 28 11. Hirunyanitiwattana, W., and S. P. Mattingly. Identifying Secondary Crash Characteristics for
29 California Highway System. Presented at 85th Annual Meeting of the Transportation
30 Research Board, Washington, D.C., 2006.
- 31 12. Zhan, C., L. Shen, M. A. Hadi, and A. Gan. Understanding the Characteristics of Secondary
32 Crashes on Freeways. Presented at 87th Annual Meeting of the Transportation Research
33 Board, Washington, D.C., 2008.
- 34 13. Zhan, C., A. Gan, and M. Hadi. Identifying Secondary Crashes and Their Contributing
35 Factors. In *Transportation Research Record: Journal of the Transportation Research Board*,
36 No. 2102, Transportation Research Board of the National Academies, Washington, D.C.,
37 2009, pp. 68–75.
- 38 14. Vlahogianni, E. I., M. G. Karlaftis, J. C. Golias, and B. M. Halkias. Freeway Operations,
39 Spatiotemporal-Incident Characteristics, and Secondary-Crash Occurrence. In *Transportation*
40 *Research Record: Journal of the Transportation Research Board*, No. 2178, Transportation
41 Research Board of the National Academies, Washington, D.C., 2010, pp. 1–9.

- 1 15. Khattak, A. J., X. Wang, H. Zhang, and M. Cetin. *Primary and Secondary Incident*
2 *Management: Predicting Durations in Real Time*. Report VTRC 87648, Virginia
3 Transportation Research Council, Charlottesville, 2009.
- 4 16. Kopitch, L., and J. Saphores. Assessing Effectiveness of Changeable Message Signs on
5 Secondary Crashes. Presented at 90th Annual Meeting of the Transportation Research Board,
6 Washington, D.C., 2011.
- 7 17. Green, E. R., J. G. Pigman, J. R. Walton, and S. McCormack. Identification of Secondary
8 Crashes and Recommended Countermeasures to Ensure More Accurate Documentation.
9 Presented at 91th Annual Meeting of the Transportation Research Board, Washington, D.C.,
10 2012.
- 11 18. Chilukuri, V., and C. Sun. Use of Dynamic Incident Progression Curve for Classifying
12 Secondary Accidents. Presented at 85th Annual Meeting of the Transportation Research
13 Board, Washington, D.C., 2006.
- 14 19. Chou, C.-S., and E. Miller-Hooks. Simulation-Based Secondary Incident Filtering Method.
15 *Journal of Transportation Engineering*, Vol. 136, No. 8, 2010, pp. 746–754.
- 16 20. Zhang, H., and A. Khattak. What Is the Role of Multiple Secondary Incidents in Traffic
17 Operations? *Journal of Transportation Engineering*, Vol. 136, No. 11, 2010, pp. 986–997.
- 18 21. Khattak, A. J., X. Wang, and H. Zhang. Spatial Analysis and Modeling of Traffic Incidents
19 for Proactive Incident Management and Strategic Planning. In *Transportation Research*
20 *Record: Journal of the Transportation Research Board*, No. 2178, Transportation Research
21 Board of the National Academies, Washington, D.C., 2010, pp. 128–137.
- 22 22. Zhang, H., and A. J. Khattak. Analysis of Cascading Incident Event Durations on Urban
23 Freeways. In *Transportation Research Record: Journal of the Transportation Research*
24 *Board*, No. 2178, Transportation Research Board of the National Academies, Washington,
25 D.C., 2010, pp. 30–39.
- 26 23. Orfanou, F. P., E. I. Vlahogianni, and M. G. Karlaftis. Detecting Secondary Accidents in
27 Freeways. Presented at Third International Conference of Road Safety and Simulation,
28 Washington, D.C., 2011.
- 29 24. Zhang, H., and A. Khattak. Spatiotemporal Patterns of Primary and Secondary Incidents on
30 Urban Freeways. In *Transportation Research Record: Journal of the Transportation*
31 *Research Board*, No. 2229, Transportation Research Board of the National Academies,
32 Washington, D.C., 2011, pp. 19–27.
- 33 25. Vlahogianni, E. I., M. G. Karlaftis, and F. P. Orfanou. Modeling the Effects of Weather and
34 Traffic on the Risk of Secondary Incidents. *Journal of Intelligent Transportation Systems*,
35 Vol. 16, No. 3, 2012, pp. 109–117.
- 36 26. Yang, H., B. Bartin, and K. Ozbay. Identifying Secondary Crashes on Freeways Using
37 Sensor Data. Presented at 92nd Annual Meeting of the Transportation Research Board,
38 Washington, D.C., 2013.
- 39 27. Yang, H., B. Bartin, and K. Ozbay. Investigating the Characteristics of Secondary Crashes on
40 Freeways. Presented at 92nd Annual Meeting of the Transportation Research Board,
41 Washington, D.C., 2013.
- 42 28. Imprialou, M. M., F. P. Orfanou, E. I. Vlahogianni, and M. G. Karlaftis. Defining
43 Spatiotemporal Influence Areas in Freeways for Secondary Accident Detection. Presented at
44 92nd Annual Meeting of the Transportation Research Board, Washington, D.C., 2013.

- 1 29. Chung, Y. Identifying Primary and Secondary Crash from Spatio-temporal Crash Impact
2 Analysis. Presented at 92nd Annual Meeting of the Transportation Research Board,
3 Washington, D.C., 2013.
- 4 30. Parker, S. T., and Y. Tao. WisTransPortal: A Wisconsin Traffic Operations Data Hub.
5 Presented at Ninth International Conference on Applications of Advanced Technology in
6 Transportation (AATT), 2006.
- 7 31. Parker, S. WisTransPortal Developer Wiki. transportal.cee.wisc.edu/resources/wiki/.
8 Accessed Jan. 7, 2013.
- 9 32. Wisconsin Department of Transportation. Official State Trunk Highway System Maps.
10 www.dot.wisconsin.gov/travel/maps/sth.htm. Accessed Jan. 7, 2013.
- 11 33. Dijkstra, E. W. A Note on Two Problems in Connexion with Graphs. *Numerische*
12 *Mathematik*, Vol. 1, 1959, pp. 269–271.
- 13 34. Hurdle, V., and B. Son. Shock Wave and Cumulative Arrival and Departure Models: Partners
14 without Conflict. In *Transportation Research Record: Journal of the Transportation*
15 *Research Board*, No. 1776, Transportation Research Board of the National Academies,
16 Washington, D.C., 2001, pp. 159–166.
- 17 35. Li, H., and R. Bertini. Comparison of Algorithms for Systematic Tracking of Patterns of
18 Traffic Congestion on Freeways in Portland, Oregon. In *Transportation Research Record:*
19 *Journal of the Transportation Research Board*, No. 2178, Transportation Research Board of
20 the National Academies, Washington, D.C., 2010, pp. 101–110.
- 21 36. Saddi, R. R. Studying the Impacts of Primary Incidents on Freeways to Identify Secondary
22 Incidents. MS thesis. University of Nevada, Las Vegas, 2009.