Can High-Resolution Detector and Signal Data Support Intersection Crash Identification and Reconstruction?

Indrajit Chatterjee *
Research Assistant, Dept. of Civil Engineering, University of Minnesota
500 Pillsbury Drive SE, Minneapolis, MN 55455, email: chat0123@umn.edu

Gary A. Davis
Professor, Dept. of Civil Engineering, University of Minnesota
500 Pillsbury Drive SE, Minneapolis, MN 55455, email: drtrips@umn.edu

Submitted to 3rd International Conference on Road Safety and Simulation, September 14-16, 2011, Indianapolis, USA

ABSTRACT

Traffic crashes may not always result in severe or fatal injuries, but can still have nontrivial impact on system performance, particularly during heavy traffic conditions. One way toward reducing the frequency of such incidents is to first identify the necessary circumstances that resulted in the collision. However, road accidents, particularly intersection related crashes, are complex phenomenon and may result from different combination of causal factors, Recently methods for recording high-resolution arterial traffic data have been developed, and it is important for traffic safety engineers to explore such high-resolution data to understand the causes of crashes. In this study we illustrate, for one particular intersection crash resulting from a signal violation, how high-resolution event-based data obtained from loop detectors can be used to identify the incident and the vehicles involved in the crash. We also illustrate how high-resolution data could support a traditional reconstruction of this crash. A Monte Carlo simulation technique was used to estimate the most probable combination of the driver behaviors that resulted in the collision. It was found that the excessive speed of the vehicle violating the red light was the most critical factor contributing to the crash.

Key words: crash identification, reconstruction, Monte Carlo, mode estimation.

INTRODUCTION

Traffic crashes may not necessarily result in fatal or severe injuries, but they can still have adverse impacts of system performance, resulting in travel delays, congestion and excessive emissions. Ultimately, reducing the frequency of crashes will require identifying their causes, and then implementing appropriate countermeasures to remove or counter those causes. Road crashes are not simple phenomena, and generally result from differing combinations of causal factors, whose relative importance can vary from case to case. For example, at an intersection a crash could have resulted from a combination of excessive speed, driver inattention, signal violation, insufficient sight distance, and poor gap choice, any one of which, if modified, could have prevented the crash. Although statistical-based safety studies can sometimes estimate aggregate causal effects, identifying how the actions of different drivers interact with each other, and with roadway features, to produce crashes requires a more microscopic approach to both the modeling of crash events and the collection of crash data. This need has in fact been recognized in the research plan for the Strategic Highway Research Program 2 (SHRP2) Safety Program, which is conducting an extensive longitudinal study using specially-instrumented vehicles, together with limited development of video-based methods for collecting data at particular sites. At present however, both the vehicle-based and site-based approaches require a temporary deployment of special equipment, and because crashes tend to be rare events, the number of actual crashes captured by such deployments will necessarily be limited. In the foreseeable future then, learning how and why crashes occur will still require post-hoc investigations of actual crash events, where the information is limited to what is available from inspection of the involved vehicles and investigation of the crash scene.

One potentially useful, but neglected, source of information is available at signalized intersections, where pavement-based sensors are used to drive traffic-actuated control logics. Pavement-based sensors register the presence of vehicles, and their output can be used to measure macroscopic variables such traffic volume, lane occupancy, and average speed. Many urban freeway systems which use pavement-based sensors to drive ramp-metering logics also archive the aggregated measures from the sensors, which provide useful data for a number of transportation-related research problems. For signalized intersections however, data archiving is more the exception than the rule and, as far as we know, archiving data at the level of the individual vehicle is not done. Recent work at the University of Minnesota has produced a package of hardware and software which, when attached to a signal controller, can record and store high resolution data for both sensor activations and signal indications. Known as SMART-SIGNAL, this package has been used to support modeling of queue formation on systems of signalized intersections, and to drive a novel method for estimating arterial travel times. In this paper we will explore the usefulness of high-resolution data made available by SMART-SIGNAL to contribute the reconstruction and causal analysis of an intersection-related crash.

To illustrate the potential value of high-resolution detector data, consider the classic problem where an estimate of a vehicle's speed is desired, and the available information consists of a measurement of skid-mark length. Letting d denote this measurement and v denote the speed of the vehicle at the beginning of the skid-mark, v is then given by

$$v = \sqrt{2 \times \mu \times g \times d}, \qquad (1)$$

where, μ is the coefficient of tire-pavement friction and g the gravitational acceleration.

If all quantities on the right-hand side of equation (1) are known with certainty then the estimated speed will also be certain, but such categorical certainty is almost never to be had. One effective way to allow for uncertainties is the Bayesian approach, where prior information regarding plausible values for v and μ are combined with an assessment of the measurement error for d to produce a posterior distribution for v. Suppose then that we have, in addition, an occupancy time from a detector with known finite measurement error. Assuming the length of the vehicle is known (if not known, it can also be estimated within the Bayesian framework), a new posterior estimate of the speed can be computed, combining the information from the observed occupancy time and the skid mark measurement. Figure 1 compares the posterior distributions of the speed computed for the two data scenarios showing the potential for detector information to improve the precision of a speed estimate.

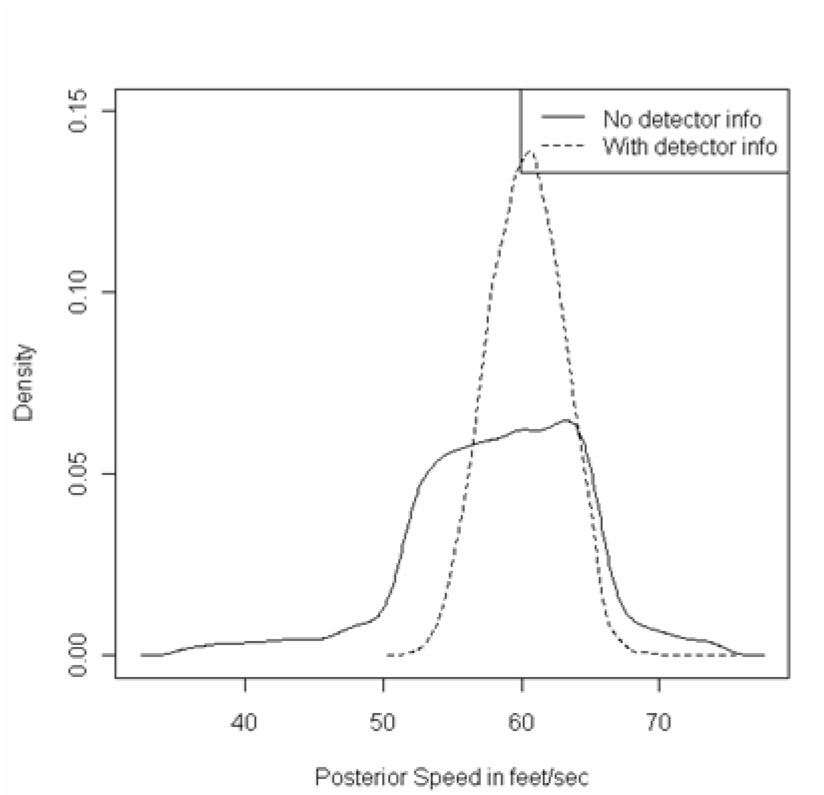

Figure 1 Posterior estimate of the speed from skid-mark and occupancy data

The rest of the paper is organized as follows: the next section briefly describes the data collected from SMART-SIGNAL system, followed by our method for matching high-resolution detector measurements to the individual vehicles involved in a crash. This will involve previous literature on incident detection on arterial networks. We then present an example where detector data are combined with more traditional crash reconstruction information, via Monte Carlo simulation, in order to compute a posterior distribution of critical driver parameters contributing to the crash, and estimate the multivariate mode as the most probable combination of parameters causing the crash.

SMART-SIGNAL DATA AND CRASH RECORDS

Recently there have been several studies based on high-resolution data to evaluate arterial performance measures such as travel time, queue length estimation and travel delay (Liu, and Ma, 2009). SMART-SIGNAL is one such integrated event based data collection and storage system. (for details see Liu and Ma, 2008). The data collection system simultaneously collects high-resolution event-based traffic data including every vehicle actuation over detectors located near intersections and every signal phase change. Two types of data can be retrieved from SMART-SIGNAL archives: vehicle arrival and departure times at the detectors and changes in the signal phase. The time stamp associated with each event enables identification of past traffic states based on the occupancy and gap measurements obtained directly from detector actuations. Currently the SMART-SIGNAL system has been installed at various intersections on Trunk Highway (TH) 55 and on Eden Prairie Centre Drive, in the Twin Cities region of Minnesota. In this paper we focus on one particular event at the intersection of TH 55 and Winnetka Ave. Figure 2 shows the layout of the intersection with detector locations and signal phase numbers.

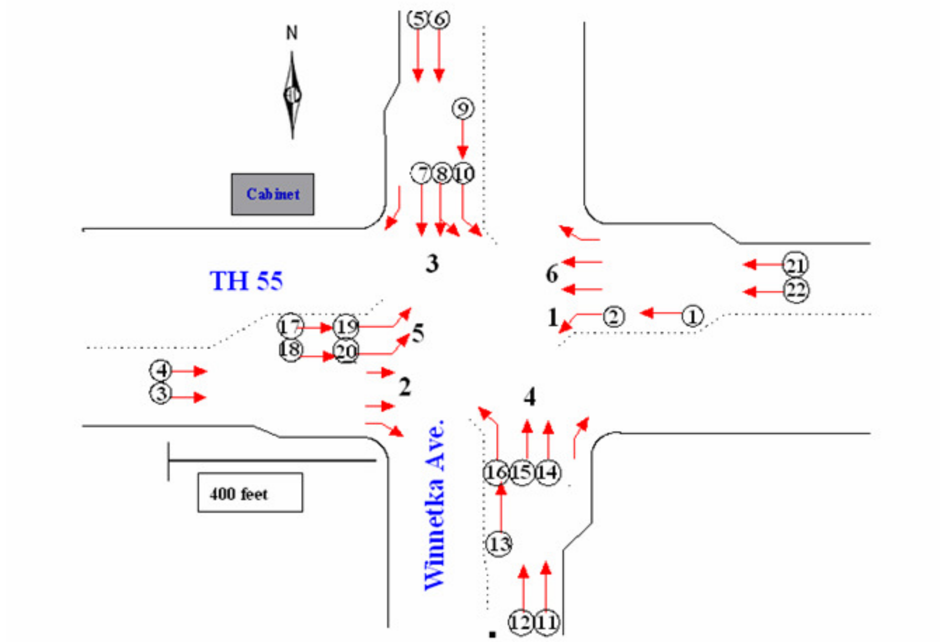

Figure 2 Smart-Signal detector layout of Winnetka Ave/ TH 55

In addition to the SMART-SIGNAL data, a crash report of the event was requested from Minnesota Department of Transportation (MnDOT). The selection of this particular event from the pool of crash records was based on our judgment that the crash was severe enough to impact the normal traffic conditions. Factors such as type of crash (in this case angle crash), location of the crash (within the intersection), time of crash (close to evening peak hour) and collision impact were considered.

Crash Description

TH 55 is an east-west running state highway, while Winnetka Ave is North-South county arterial. TH 55 EB (approaching the intersection) has two through and two left turn lanes with a separated right turn, while Winnetka south bound has two through lanes of which one is shared left turn lane, and a separate left-turn and right-turn lane. The following section gives the description of the event as documented in the police crash report.

According to a witness, unit 1 was driving eastbound on TH 55, while unit 2 was southbound on Winnetka and approaching the intersection with a green light. Unit 1 (a 1995 Chevy Blazer) violated the signal and collided with unit 2 (a 2002 Buick LeSabre). The collision then caused unit 1 to hit a NB vehicle which was waiting at a red light. The time of the crash was recorded as 16:07 in the report. The description of the event suggests that the crash was severe enough to impact traffic conditions at the intersection. Also information about the vehicle type could help to calculate speeds of the vehicles involved in the crash based on occupancy data from the detectors.

CRASH IDENTIFICATION PROBLEM

The main idea behind this study is based on the observation that severe crashes tend to disrupt normal traffic flows and so it may be possible using archived event-based data (including both vehicle-detector actuation events and signal phase change events) from SMART-SIGNAL, to first determine when the crash occurred, and then identify the corresponding signal indications and the loop detector data for the involved vehicles. The crash identification problem can be treated as a type of Incident Detection Problem. Although we are not proposing a new algorithm for incident detection it will be a helpful to review some of the relevant literature on incident detection on arterial networks.

Literature Review

In the past decade or so there has been a constant effort to improve the efficiency of existing roadways. One of the means to accomplish this goal is the introduction of incident management systems that involve initial incident detection, emergency vehicle co-ordination, clearing the incident site, informing motorists and restoring normal traffic conditions. A substantial literature is devoted to freeway incident detection, and somewhat less to signalized urban arterial incident detection, but a common feature associated with incidents is a sudden decrease in roadway capacity due to lane blockage, leading to reduction in speeds and longer occupancy times at detector stations. However, to be detected an incident must be severe enough for traffic data to reflect a noticeable deviation from normal conditions (Culip and Hall, 1997).
Incident detection algorithms fall broadly into two categories: (1) Pattern recognition algorithms which compare current traffic patterns with historic traffic data (the most common source of data being loop detectors), where a difference exceeding some predefined threshold value (calibrated based on the location) indicates an incident (Han and May, 1989; Solomon, 1991). However such algorithms may not be able to differentiate between incidents and recurring congestion. (2) Short term prediction algorithms that involve statistical procedures such as time series to predict future traffic measurements, and an incident is detected if the current measurements fall outside the confidence limits of the forecast (Ahmed and Cook, 1982; Bell and Thancanamootoo, 1988). Han and May (1988) indicated that an arterial incident detection algorithm based on artificial intelligence was more likely to be effective for incidents which are in proximity to detector locations. A similar fact was mentioned in Sethi et al. (1995), where the authors concluded that a higher detection rate was achieved for incidents locating near or at intersections, because such incidents are expected to have greater impact on traffic conditions.

Interestingly, a seemingly unrelated area receiving attention is land cover change detection for earth science. Generally such studies involve satellite image comparisons based on pixels, as well as traditional time series approaches to detect land changes (Lunetta et al., 2006). Change detection problems involving time series data deal with two separate problems (a) detection (whether change has occurred or not), (b) estimation (location of change point). A wide variety of literature based on both frequentist and Bayesian techniques for change detection is available (Krishnaiah et al., 1988; Chen and Gupta, 2001). Change detection techniques originated in manufacturing process for quality control, where time series measurements were expected to follow a certain underlying distribution and any deviation was identified as an error in the manufacturing process. Such a method is often termed as a parameter change approach, where the generating process is characterized by a parameter of the underlying distribution. A substantial literature is available for change detection based on such techniques (Chen and Gupta, 2001). Often, with parameter change technique an additional test statistic is used to determine if a change point exists or not.

One such popular parameter change technique was introduced by Page (1957) based on a statistic called cumulative sum (CUSUM) to detect a changes in a mean value. A wide application of CUSUM can be found in many change detection studies (Chernoff and Zacks, 1964; Jandhyala et al., 2002), and here we will use the CUSUM statistic to detect changes in detector occupancy time. The basic idea behind CUSUM is based on prior knowledge of an expected measure of a process. CUSUM then calculates the cumulative deviation of the current measure from its expected value. For example let $\{t_1, t_2, \ldots, t_n\}$ be a time series of observations with a known expected value, say μ. Then CUSUM statistic, CS_k at k time point is calculated as

$$CS_k = \sum_{i=1}^{k} (t_i - \mu) \qquad (2)$$

Identification Methodology

In this study, once the time of the event was approximately located from the crash report archived detector and signal data from SMART-SIGNAL system were obtained for a time window bracketing the crash occurrence (half an hour before and after the incident as documented in the crash report). Then a local query was set using Excel to segregate the detector actuation events (i.e. occupancy time) based on whether the corresponding signal phase was red or green. Now recall that Figure 2 shows the detector layout of the intersection. Sequences of occupancy plots corresponding to separate green and red phases for detectors 4 and 3, located on east bound TH 55 about 400 feet upstream of the intersection, were prepared. Figure 4 shows the occupancy time measurements around 16:07 for detector 4 when the signal phase was red for the approach. It is evident from the figure that around 16:06 an event with very low occupancy time (i.e. high speed) was observed. For comparison, the general occupancy pattern during the red phase for the same approach was evaluated for an adjacent period of the incident (see Figure 3).

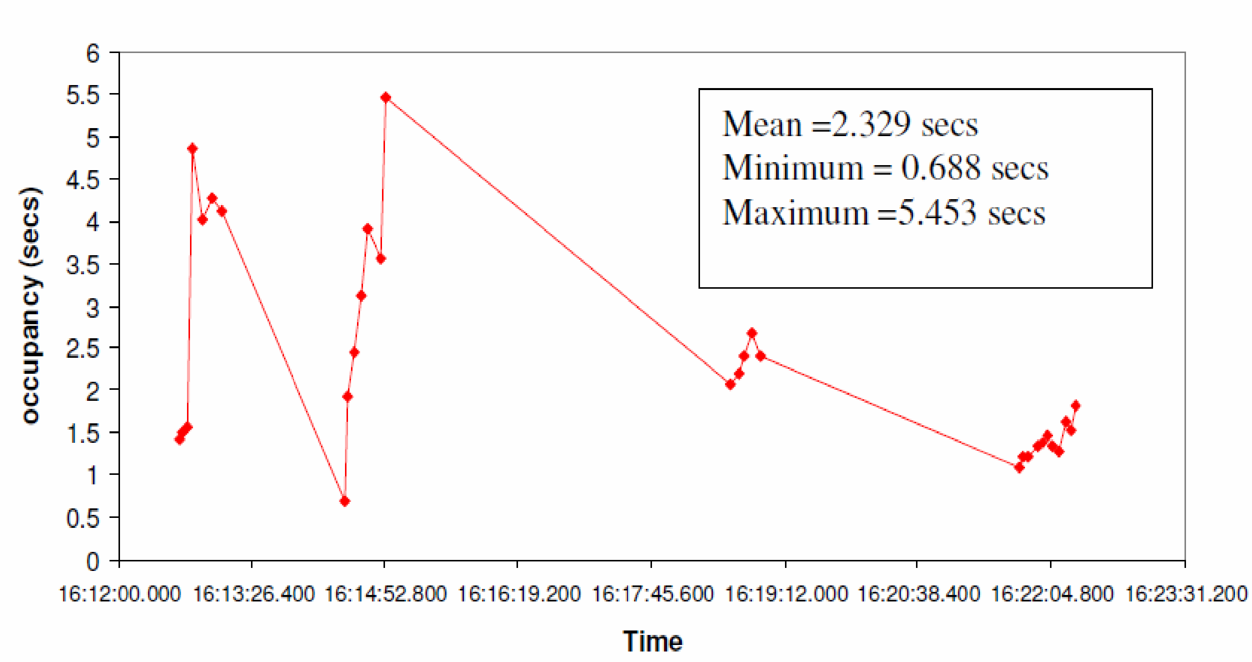

Figure 3 General occupancy pattern for detector 4 (red phase)

Descriptive statistics, with mean=2.329 secs, minimum=0.688 secs and maximum=5.453 secs, for general occupancy pattern suggested that the occupancy of 0.235 secs (from Figure 4) is below the general pattern, which indicates a potential target vehicle with much higher speed in comparison to other vehicles approaching the intersection during the red phase. However, at this point we might not be able to identify with certainty this vehicle as the vehicle ultimately involved in the crash. For this we examined the possibility of the immediate recorded vehicles around the target vehicle based using the CUSUM statistic.

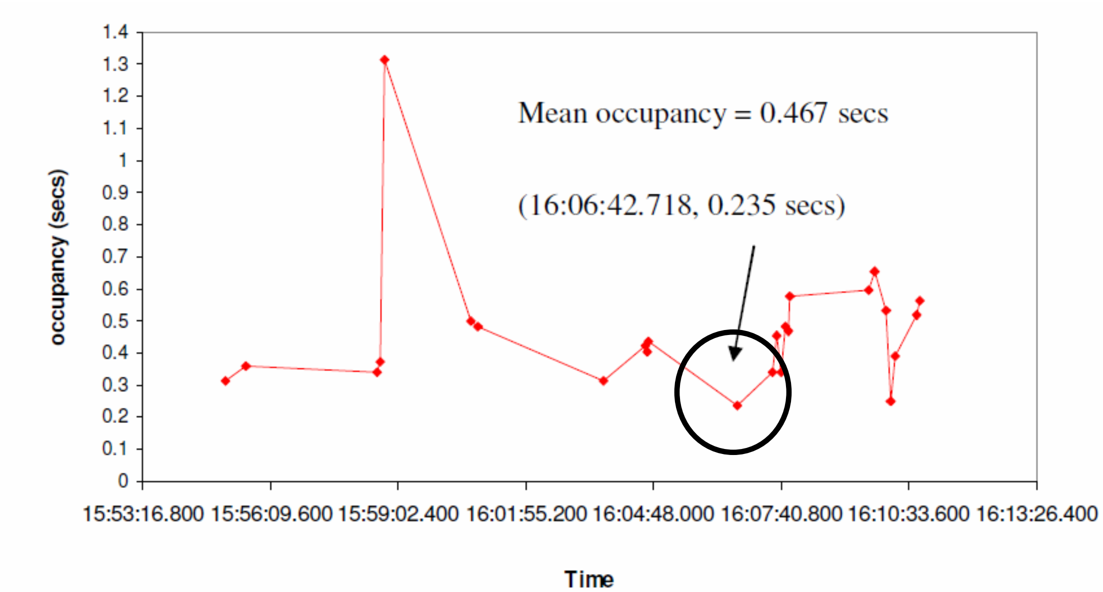

Figure 4 Occupancy data for detector 4 during red phase and identifying vehicle 1

To extract further information about the arrival of the target vehicle at detector station 4 it was found that when the vehicle arrived at the detector, the signal for the phase had been red for 3.698 secs. We looked at similar occupancy plots for detector station 3 during the red phase and could not find any potential vehicle with a similar low occupancy time. Next, an effort was made to determine if a change in traffic conditions could be detected after the incident. For this purpose, the occupancy of the detectors 3 and 4 corresponding to the green phase approximately around the time of crash was investigated. Our hypothesis states, given that a congestion-inducing incident happened, the traffic states (in our study, occupancy) would be significantly different from the traffic states under normal traffic conditions. The occupancy pattern for detector 4 during the green phase around the reported crash time, i.e., 16:07 was plotted, as shown in Figure 5. The plots clearly indicate a distinct change in the pattern of occupancies at approximately 16:06. The mean green phase occupancies before 16:06 was found to be 0.343 secs compared to the much higher average high occupancy of 0.782 secs after 16:06. Figure 6 shows a similar pattern with a gradual increase in occupancy at detector 3 during the green phase following the incident.

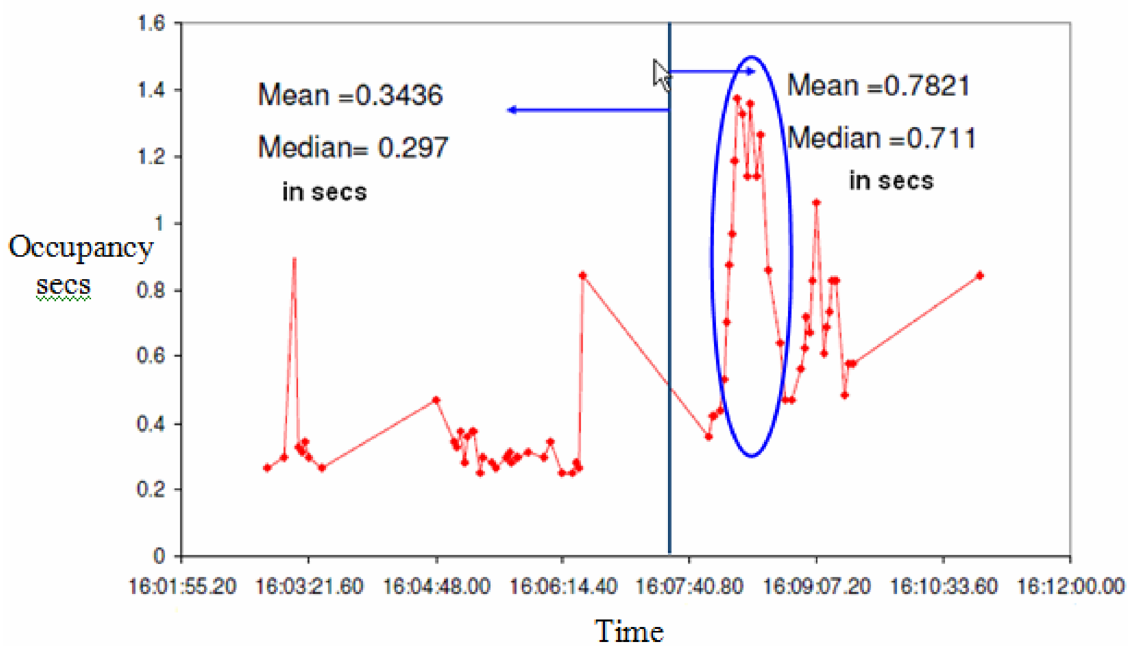

Figure 5 Post-incident occupancy pattern at detector 4 during green phase

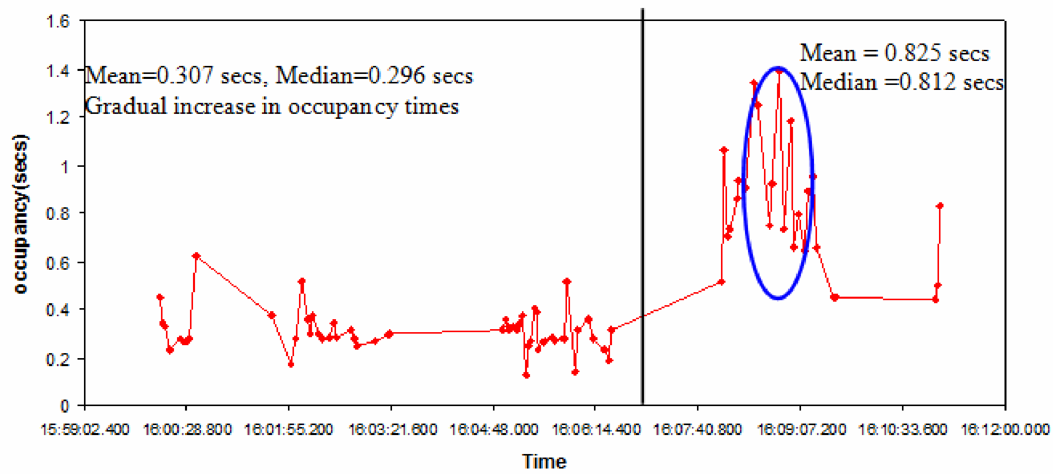

Figure 6 Post-incident occupancy pattern at detector 3 during green phase

The CUSUM statistic, introduced in the previous section, was used to measure any deviation from the general pattern in lane occupancies before and after the incident, during the green phase. Figure 7 shows the CUSUM plot along with the occupancy for detector 3. The figure detects a change in the occupancy trend after 16:06:27.078.

The initial variation in the CUSUM statistic is due to the randomness usually observed in any arterial traffic system. The prior information about the mean occupancy was calculated based on occupancy data from previous time periods under normal traffic conditions. One of the interesting facts about CUSUM, which can be observed from the above plot, is that once the process deviates from its general trend due to the incident, CUSUM increases very rapidly and also monotonically. This monotonic behavior of CUSUM is quite effective from a detection algorithm point of view. For example, if we want to write an algorithm based on the CUSUM statistic to detect on incident, the only thing we have to find is the point where this monotonic behavior begins, which is relatively easy to implement.

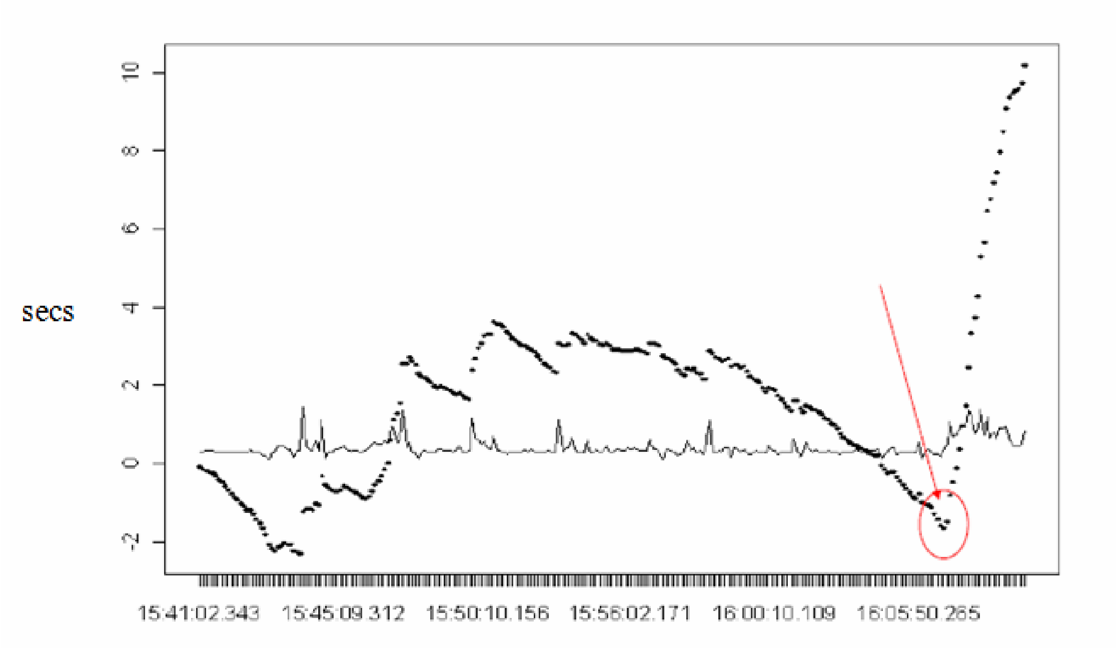

Figure 7 Cusum statistic for incident detection from occupancy data at detector 3

The CUSUM statistic plot as shown in Figure 7 suggested the change in the pattern of the occupancy during green time had occurred after 16:06. And when we looked at Figure 3, the occupancy of the vehicle just before the potential target vehicle (with occupancy 0.235 secs) was 0.438 at 16:04:42. Since the effective length of the vehicle is known from the preliminary crash report (21.68 feet), the point estimate of the speed of the vehicle could be computed as 49.49 feet/sec and hence consequently the estimated time to arrival to the potential conflict point (which is about 435 odd feet away from the detector location) would be 16:04:42+ (435.34/49.49 secs) =16:04:50.79, which clearly does not conform with the CUSUM result. Also, if we look at the occupancy data observed for the vehicle just recorded following the identified vehicle, we can see from Figure 4 that the occupancy was measured as 0.343 secs at 16:07:29.953. Similarly, the estimated speed of the vehicle would be 63.20 feet/sec and the estimated time to arrive at the potential conflict point is 16:07:36.84. However, a very high occupancy time (around 170 secs) for detector 8 with green phase on SB Winnetka Ave (recall Figure 2), was recorded at 16:06:52.906, indicating an incident occurring before that time. Hence based on the results from the CUSUM statistics, we have a bound for time of crash between 16:06:27.078 and 16:06:52.906. Therefore, the most probable vehicle which was involved in the crash from Figure 4 was identified with occupancy of 0.235 secs at 16:06:42.718.

One key point to be noted here is that, although we were able to identify a change in the general occupancy pattern based on CUSUM statistic, it is not certain at this point whether this change in the pattern is due to incident or to some recurrent congestion, such as spill back from a downstream link. To distinguish between incident congestion and recurrent congestion it was decided to check the occupancy pattern of the main-line (i.e., TH 55 EB) detectors at the downstream link, located approximately 460 feet downstream of the crash location around 16:06, as suggested in the study by Gall and Hall (1989). If a low occupancy pattern is found, indicating high speed, this would suggest no spillback phenomenon and hence the congestion at the upstream link could be attributed to the incident. The observed low mean occupancy time for the downstream detectors suggested that there was no spillback phenomenon. Hence the congestion observed upstream at Winnetka Ave can be attributed to the incident.

The next step is to identify unit 2, the vehicle which was SB on Winnetka Ave with a green phase. The crash report indicated that unit 2 was a 2002 Buick LeSabre. The speed estimate from a single detector is given by

$$\text{Speed} = \text{Effective length/ Occupancy time}, \qquad (4)$$

where, Effective length = sum of the vehicle and detector length. For unit 2 effective length = (16.6+6) feet, where length of the detector is 6 feet.

The point of collision based on the approach directions of the two vehicles was found be 123.86 feet from detector stations 7 and 8. Occupancy data for detectors 7 and 8 were extracted corresponding to the green phase around 16:07, the time of collision occurrence as recorded in the crash report. Unit 2 was then identified as the most probable vehicle recorded at detector 7 or 8 with an estimated speed, based on the occupancy time, to arrive at the collision point before the upper bound identified above. That is, unit 2 was the vehicle recorded at detector 7 at 16:06:45.937 with occupancy time of 1.438 secs. Hence an estimated arrival time of unit 2 at the collision point is 16:06:45.937 + (123.86/22.6/1.438) = 16:06:53.818. This estimated arrival time for unit 2 was calculated based on constant speed; however the signal phase data suggest that the unit 2 was discharging during green phase and may have some acceleration and so arrived at the collision point before the estimated time, but within the bound established above.

CRASH RECONSTRUCTION

Drawing from Baker's (1990) notion of crash reconstruction of determining how a crash occurred, we next tried to address a question that, given the initial speed estimates and locations (both space and time) from detector data, what could be learned about the behavior of the drivers involved in the crash. To answer this, trajectories of the two approaching vehicles were modeled by numerically solving a system of ordinary differential equations. Vehicles were initially assumed to travel at uniform speed along straight lines.

Given initial speed and location, at each time step of 0.01 secs the vehicle's speed and locations were updated using a simple Euler's formula, and each time the separation distance (distance between the centre of two vehicles), d was calculated to check whether two vehicles had collided or not. Figure 8 shows the plot of separation distance between the two vehicles assuming constant speeds, and it is clear that had the drivers not taken any action, the crash would not have occurred. Also, it was observed from the simulation that unit 1 arrived at the collision point much earlier than unit 2, suggesting that for crash to happen, unit 1 had to decelerate and/or unit 2 had to accelerate.

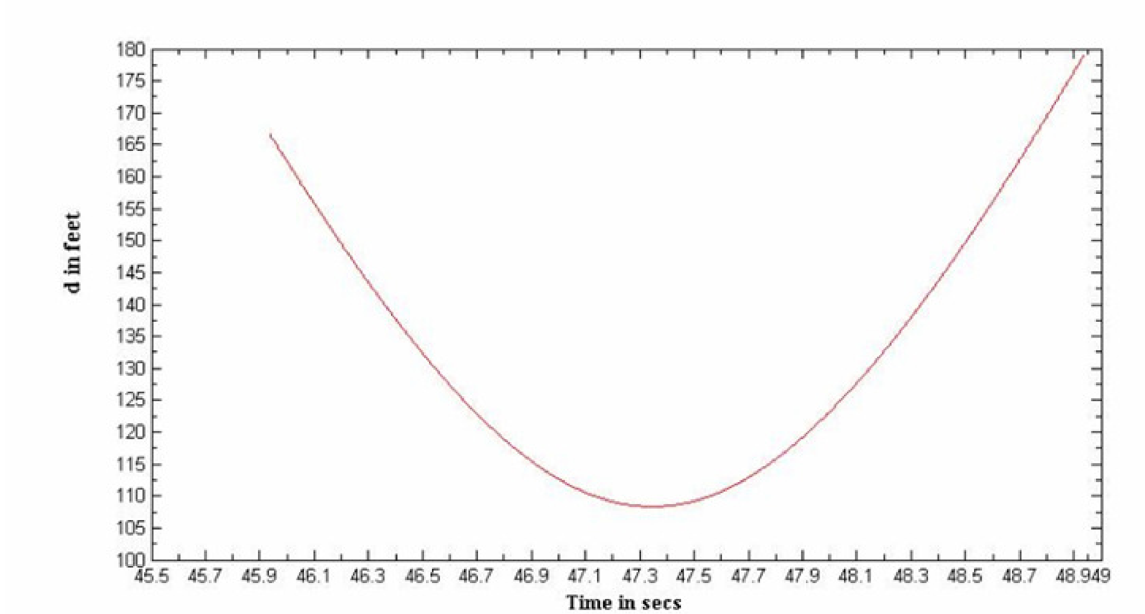

FIGURE 8 Separation Distance Based on Initial Speed Estimates.

Traditionally, accident reconstruction involves estimation of the speed of the vehicles at the point of collision from post collision information such as final resting positions for the vehicles, skid marks and damage. Such detailed information was not available for this particular case. Hence here, we would like to illustrate how, given such post collision details, it is possible to estimate driver behaviors contributing to the crash. For this purpose, a plausible hypothetical post collision scenario was added, where the following assumptions on the final position of the vehicles were made.

1. The crash report suggested that unit 1, after colliding with unit 2, rolled down to eventually collide with unit 3 which was stopped at red light NB Winnetka (approximately 60 feet from collision point). Based on the report it was assumed that subsequent impact with unit 3 was minimal, i.e., unit 1 came to a complete stop just at the point of collision with unit 3.
2. Unit 2 was assumed to skid after it collided with unit 1 with a post-impact skid mark suggesting that unit 2 came to a complete stop somewhere between 5 to 20 feet from collision point, with an angle of departure (Φ) for unit 2 as 30^{o}. Figure 9 illustrates the collision scenario.

The post impact speeds of the two vehicles were evaluated based on the stopping distance formula,

$$V_i = \sqrt{2 \times f_i \times g \times s_i}, \quad \text{where } g = 32 \text{ feet/sec}^2, \ i = 1, 2 \qquad (5)$$

For unit 1, f_1 (rolling resistance) was assumed to be 0.15 and $s_1 = 60$ feet
and for unit 2, f_2 (skidding resistance) was assumed to be 0.7 and s_2 between 5 and 20 feet,

$$\text{i.e., } \sqrt{2 \times 0.7 \times 32 \times 5} < V_2 < \sqrt{2 \times 0.7 \times 32 \times 20} \quad \text{(in feet/sec)} \qquad (6)$$

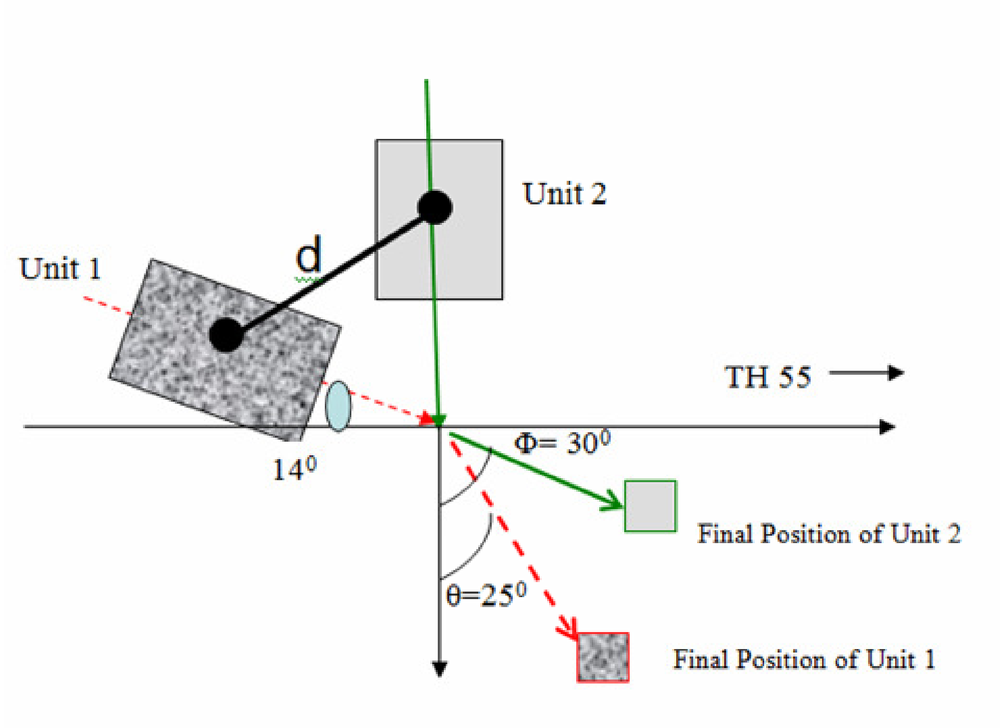

Figure 9 Collision diagram under hypothetical scenario

The next step was to evaluate the pre-impact speeds of the vehicles based on point-mass collision theory (Brach and Brach, 2004). Let v_1 *and* v_2 denote the pre-impact speed, and m_1 and m_2 be the corresponding masses of unit 1 and unit 2 respectively. Then the conservation of momentum equation in the x and y directions are given by,

$$m_1 \times v_1 \times Cos(14^\circ) = m_1 \times V_1 \times Cos(65^\circ) + m_2 \times V_2 \times Cos(60^\circ) \quad (7)$$

$$m_1 \times v_1 \times Sin(14^\circ) + m_2 \times v_2 = m_1 \times V_1 \times Cos(25^\circ) + m_2 \times V_2 \times Cos(30^\circ) \quad (8)$$

Substituting the values for V_1 and V_2 into (7) and (8), bounds for v_1 and v_2 were established as

$$16.245 < v_1 < 24 \quad and \quad 26.7 < v_2 < 39 \quad (9)$$

Finally, the collision set was established as

$$\{(d, v_1, v_2) : d < d_{crit}, 16.245 < v_1 < 24.0 \text{ and } 26.7 < v_2 < 39.0\}, \quad (10)$$

where, d_{crit} is the crash closeness threshold.

Perhaps one input that needs more explanation is the collision closeness threshold (d_{crit}) which was set to be 14.7 feet in the simulation model based on the dimension of the vehicles involved in the crash and their angle of approach (76^o). The simulation model used this threshold value to decide whether a crash has occurred or not and hence it is desirable to make sure that the value was reasonable.

A 3-parameter model was assumed for this study, (a) unit 2 acceleration (*acc2* in feet/sec^2), (b) unit 1 deceleration (*dcc1* in feet/sec^2) and (c) for driver 1, perception-reaction time (*rt* in secs) which was defined as the time elapsed between unit 1 arriving at the detector 4 and the initiation of braking. Our objective was to find the joint distribution of the three parameters (*acc2, dcc1* and *rt)* conditioning on the crash occurrence. A Monte Carlo simulation technique based on rejection sampling was adopted to sample the desired distribution (Robert and Casella, 2004). The initial distribution for *acc2* was chosen as Uniform between 0 to 10 feet/sec^2, where 10 feet/sec^2 was deemed to be a reasonable upper bound for acceleration. *dcc1* was sampled from Uniform between -5 feet/sec^2 and -21 feet/sec^2. The upper bound for *dcc1* was chosen from the fact that for the given initial speed estimates and location of the vehicles any deceleration weaker than -5 feet/sec^2 would not result in a collision. The lower bound was chosen based on the study by Fambro et al. (1997), where the mean emergency braking rate was taken as -.65g. For the perception-reaction time of unit 1, initial samples were drawn from Uniform [0, (45.937-42.718)]. The upper bound was chosen on the basis of the fact that if unit 1 had a perception-reaction time longer than 3.219 secs, the crash would not have occurred even for the strongest braking rate. Once the three parameters were sampled, for each time step, the location and speed of the two vehicles were updated using the simple Euler's method as mentioned before, and the separation distance (*d* in feet) was computed. If the condition for crash in equation 10 was satisfied then a collision was recorded. Monte Carlo Simulation was computed in R (2008) statistical software. Table 1 shows the descriptive statistics of the parameters obtained from the simulation. The point estimate of the collision time from the simulation was found to be 16:06:50.481, which is within the bound suggested by CUSUM statistics.

TABLE 1 Sample Statistics from Monte Carlo Simulation

Parameters	**Mean**	**Standard deviation**
acc2(feet/sec^2)	4.65	0.529
dcc1(feet/sec^2)	-10.157	0.632
rt (secs)	0.337	0.251

Mode Estimation

Traffic accident reconstruction can be treated as an effort to determine how an accident has occurred, and often involves uncertainties regarding the behavior of the drivers involved in the crash. One way to identify a probable cause is to estimate the mode of the posterior distribution of the parameters obtained from Monte Carlo simulation. The mode can be defined as the most probable combination of factors that resulted in the particular collision.

The study of mode estimation from a sample can be found in Parzen (1962), where the author defined a family of kernel estimators for a distribution *f(x)* and the mode estimates were shown to be asymptotically normal and consistent. Later on, Silverman (1986), Wand and Jones (1993) explored more in this area of kernel density estimation. Initially the focus was on univariate distributions but later the theory was extended to the multivariate case. The next section gives a brief introduction to Kernel density estimation.

Kernel density estimation

Suppose a s-variate random sample of size n, $X_1, X_2, X_3, \ldots\ldots, X_n$ is drawn from an unknown distribution f. Then the Kernel density estimate of f is given by,

$$\hat{f}(x:H) = n^{-1}\sum_{i=1}^{n} K_H(x - X_i), \tag{11}$$

where, $x = (x_1, x_2, \ldots\ldots\ldots, x_s)^T$. If $K(x)$ is the kernel, which is a probability density and H is bandwidth matrix, which is a symmetric and positive definite matrix, then

$$K_H(x) = |H|^{-1/2}\ K(H^{-1/2}x) \tag{12}$$

The choice of K is not as important as the choice of bandwidth matrix, as the amount of kernel smoothing is controlled by the bandwidth matrix (Wand and Jones, 1994). A substantial literature is available on optimal bandwidth selection (Duong and Hazelton, 2003). Here, without going into the details we highlight the key notion underlying the bandwidth selection procedure. The form of $K(x)$ is taken as standard normal in this study, i.e.

$$K(x) = (2\Pi)^{-s/2} \exp(-\frac{1}{2} x^T x) \tag{13}$$

The common measure of performance of $\hat{f}$ used in the literature is Mean Integrated Squared Error (MISE),

$$MISE(H) = E \int_{R^s} \{\hat{f}(x,H) - f(x)\}^2 dx \tag{14}$$

The optimal bandwidth is given by

$$H_{opt} = \arg\min_{H} MISE(H) \tag{15}$$

over all possible symmetric positive definite s by s matrices.

Two separate optimal bandwidth selectors were used in this study: A Plug-in bandwidth selector (with pre-sphering of data) and a Least squares cross validation based on leave-one-out estimator. Details are given in the literature mentioned above. Bandwidth results from the two selectors when compared were found similar, so the H matrix evaluated by Plug-in bandwidth selector (based on sample covariance matrix) was used for further analysis. The "Ks" package (Duong, 2007) for multivariate kernel smoothing was used in the R statistical language.

Evaluating the mode of a multivariate distribution is itself a challenging task. However, some progress has been made in the past (Devroye, 1979; Tsybakov, 1990). One of the simple techniques, adopted in this study, was recommended by Abraham et al. (2003). The mode is estimated by maximizing the kernel estimate over the set of sample values. For large enough samples and under certain mild conditions this estimate is found to be consistent and converges almost surely to the true value. More formally, the estimated mode is defined as follows.

If S_n denotes the set of sample points $\{X_1, X_2, \ldots, X_n\}$ then the estimated mode α_n is defined as

$$\alpha_n \in \{x \in S_n : f_n(x) = \max_{1 \le i \le n} f_n(X_i)\}, \quad (16)$$

where, $f_n(X_i)$ is the Kernel density evaluated at sample point X_i.

To compare with Abraham's estimate for mode, an optimization routine in R is called, which returns a vector {*acc2*, *dcc1* and *rt*} that maximizes the Gaussian kernel density estimate (based on the evaluated bandwidth matrix) according to equation 11. Abraham's estimates are found to be {*acc2*=4.935 feet/sec^2, *dcc1*= -9.978 feet/sec^2, *rt*= 0.262 secs}, where the mode estimates using optimization routine were {*acc2*=4.76 feet/sec^2, *dcc1*= -9.94 feet/sec^2, *rt*= 0.245 secs}

CONCLUSIONS AND SUMMARY

Recent traffic studies have witnessed increased use of high-resolution arterial traffic data to evaluate various traffic performance measures. It is also important for traffic safety engineers to explore such high resolution data to improve traffic safety. This study demonstrates how one such integrated event-based data collection and storage system, developed at University of Minnesota, along with a crash report could be used to identify features of an intersection crash involving a signal violation. When post collision information is available, these data can help to reconstruct the event and identify the most probable reason for the crash. It was fairly straight forward to identify the crash based on occupancy data extracted from the SMART-SIGNAL system. The CUSUM statistic was used to detect the change in the pattern of occupancies before and after the incident. It was found quite effective and easy to implement, and the suggested bound for time of collision was between 16:06:27.078 and 16:06:52.906 (i.e. within 25.83 secs). However, one important point to note that this methodology may not yield good results for all cases, particularly for less severe incidents (in terms of traffic impact) or events during very light traffic conditions.

Once the crash was identified, the next effort was to reconstruct the event. The detector data and the crash report were not sufficient for this, so a plausible hypothetical scenario was developed where the final position of the vehicles was assumed, in order to illustrate how the driver behaviors involved in the crash could be estimated given the post-collision knowledge. The initial speed estimates for the two vehicles were obtained based on the knowledge of vehicle type recorded in the crash report and the occupancy data from the detector. A 3-parameter model was developed and a Monte Carlo simulation technique was used to estimate the conditional distribution of the parameters given the crash had occurred. The estimated mode of the multivariate distribution based on the two methods were quite close, and results indicated that even with a relatively strong braking effect the crash could not be avoided, suggesting that the excessive speed of the unit 1 was a significant factor for the signal violation and consequently the crash occurrence. The reconstruction effort indicated that it would be interesting to investigate scenarios such as fatal crashes, where more specific details regarding post-collision status of the vehicles are available from detailed crash investigations.

ACKNOWLEDGEMENT

The authors wish to express their warmest appreciation to Dr. Henry Liu and his students for providing the relevant SMART SIGNAL data and detector layout map for the intersection. This research was supported by the Intelligent Transportation Systems Institute at the University of Minnesota. However, the authors are solely responsible for the content of this paper.

REFERENCES

Abraham, C., G., Biau and B. Cadre (2003). Simple estimation of the mode of a multivariate density. *The Canadian Journal of Statistics*, Vol. 31, No. 1, pp 23-34.

Ahmed, S. A. and A. R. Cook (1982). Application of Time-Series Analysis techniques to Freeway Incident Detection. In *Transportation Research Record 841*, TRB, National research Council, Washington, D.C., pp. 19-21.

Baker, J. S. (1990). Causes and contributing factors in traffic accidents. Traffic Accident Reconstruction. Northwestern University Traffic Institute.

Bell, M. G. H., and B. Thancanamootoo (1988). Automatic Incident Detection Within Urban Traffic Control Systems. *Proc., Roads and Traffic 2000 Conference*, Vol. 4(2), Sept.

Boriah, S. (2010). *Time Series Change Detection: Algorithms for Land Cover Change*. PhD dissertation, Department of Computer Science and Engineering, University of Minnesota,

Brach. R.M., and R.M. Brach (2004). *Vehicle Accident Analysis and Reconstruction Methods*, SAE International, Pennsylvania,

Chen, J. and A. Gupta (2001). On change point detection and estimation. *Communications in Statistics: Simulation & Computation,* 30(3), 665-697,

Chernoff, H. and S. Zacks (1964). Estimating the current mean of a normal distribution which is subjected to changes in time. The Annals of Mathematical Statistics, 35 (3), 999-1018.

Culip, M.J. and F. Hall (1997). Incident Detection on an Arterial Roadway. In *Transportation Research Record 1603*, TRB, National research Council, Washington, D.C., pp. 112-118.

Devroye, L. (1979). Recursive estimation of the mode of a multivariate density. *The Canadian Journal of Statistics*, 7, 159-167.

Duong, T., M.L. Hazelton (2003). Plugin bandwidth matrices for bivariate kernel density estimation. Journal of Nonparametric Statistics, 15, pp. 17-30.

Duong, T. (2007). ks: Kernel Density Estimation and Kernel Discriminant Analysis for Multivariate Data in R. *Journal of Statistical Software*, Volume 21, Issue 7.

Fambro, D., Fitzpatrick, K., and kappa, R. (1997). *NCHRP Report 400: Determination of Stopping Sight Distance.* TRB, National Research council, Washington D.C.

Gall, A.I., F. Hall (1989). Distinguishing between incident congestion and recurrent congestion: A proposed Logic. In *Transportation Research Record 1232*, TRB, National research Council, Washington, D.C., pp. 1-8.

Han, L. D., and A. D. May (1988). Artificial Intelligence approaches for signalized network control. Research Report UCB-ITS-RR-89-15. Institute of Transportation Studies, University of California at Berkeley.

Han, L. D., and A. D. May (1989). *Automatic Detection of Traffic Operational Problems on Urban Arterials*. Research Report UCB-ITS-RR- 89-15. Institute of Transportation Studies, University of California at Berkeley.

Jandhyala, V. K., S. B. Fotopoulos, and D. M. Hawkins (2002). Detection and estimation of abrupt changes in the variability of a process. *Computational Statistics & Data Analysis*, 40(1), 1-19.

Krishnaiah, P. R. and B. Q. Miao (1988). Review about estimation of change points. In Quality Control and Reliability, Volume 7 of Handbook of Statistics, pp. 375-402. North-Holland.

Liu, H., and W. Ma (2008). A Real-time Performance Measurement System for Arterial Traffic Signals. Presented at 86th Annual Transportation Research Board meeting, Washington, D. C.

Liu, H., and W., Ma (2009). A virtual probe approach for time-dependent arterial travel time estimation. *Transportation Research, Part C* 17 (1), 11–26.

Lunetta, R. S., J. F. Knight, J. Ediriwickrema, J. G. Lyon, and L. D. Worthy (2006). Land-cover change detection using multi-temporal MODIS NDVI data. Remote Sensing of Environment, 105(2):142-154.

Page, E. S (1957). On problems in which a change in a parameter occurs at an unknown point. *Biometrika*, 44(1/2), 248-252.

Parzen, E (1962). On estimation of a probability density function and mode. The Annals of Mathematica1Statistics, 33, 1065-1076.

Robert, C.P. and Casella, G. (2004). Monte Carlo Statistical Methods (second edition). New York: Springer-Verlag.

R (2008): A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org

Sethi, V., N. Bhandari, F. S. Koppelman, and J. L. Schofer (1995). Arterial Incident Detection Using Fixed Detector and Probe Vehicle Data. *Transportation Research C*, Vol. 3C, No. 2, pp. 99–112, April.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

Solomon, M. (1991). *A Review of Automatic Incident Detection Techniques*. ADVANCE Program Technical Report NU-1d.1-1, Transportation Center, Northwestern University.

Tsybakov A (1990). Recursive estimation of the mode of a multivariate distribution. *Problems of Information Transmission*, 26, 31-37.

Wand, M.P and M.C. Jones (1993). Comparison of Smoothing Parameterizations in Bivariate Kernel Density Estimation. *Journal of American Statistical Association*, 88, 520-528.

Wand, M.P and M.C. Jones (1994). Multivariate Plug-in Bandwidth Selection. *Computational Statistics*, 9, 97-116.