Title
Dynamic All-Red Extension at Signalized Intersection: Probabilistic Modeling and Algorithm

Authors
Zhang, Liping
Zhou, Kun
Zhang, Wei-bin
et al.

Publication Date
2011

CALIFORNIA PATH PROGRAM
INSTITUTE OF TRANSPORTATION STUDIES
UNIVERSITY OF CALIFORNIA, BERKELEY

Dynamic All-Red Extension at Signalized Intersection: Probabilistic Modeling and Algorithm

Liping ZHANG, Kun Zhou, Wei-bin Zhang and James A. Misener

California PATH Working Paper
UCB-ITS-PWP-2011-01

This work was performed as part of the California PATH Program of the University of California, in cooperation with the State of California Business, Transportation, and Housing Agency, Department of Transportation, and the United States Department of Transportation, Federal Highway Administration.

Report for Caltrans task order 65A0363

January 2011

ISSN 1055-1425

Dynamic All-Red Extension at Signalized Intersection: Probabilistic Modeling and Algorithm

Liping ZHANG, Kun Zhou, Wei-bin Zhang and James A. Misener

California PATH, University of California, Berkeley
1357 South 46th st,
Richmond, CA 94804,
{lpzhang, kzhou, wbzhang, misener}@path.berkeley.edu

Acknowledgment

This work was sponsored by the State of California Business, Transportation and Housing Agency, Department of Transportation (Caltrans) and U.S. Department of Transportation (DOT) as part of the CICAS project (Caltrans task order 65A0363). The contents of this paper reflect the views of the authors, who are responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of California.

DISCLAIMER STATEMENT

This document is disseminated in the interest of information exchange. The contents of this report reflect the views of the authors who are responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of California or the Federal Highway Administration. This publication does not constitute a standard, specification or regulation. This report does not constitute an endorsement by the Department of any product described herein.

For individuals with sensory disabilities, this document is available in Braille, large print, audiocassette, or compact disk. To obtain a copy of this document in one of these alternate formats, please contact: the Division of Research and Innovation, MS-83, California Department of Transportation, P.O. Box 942873, Sacramento, CA 94273-0001.

Abstract

Dynamic All-Red Extension (DARE) has recently attracted research interest as a non-traditional intersection collision avoidance method. We propose a probabilistic model to predict red light running (RLR) hazard for dynamic all-red extension system. The RLR hazard, quantified by a predicted encroachment time, has contributory factors including the speed, distance and car-following status of the violator and the empirical distribution of the entry time of conflict traffic. An offline data analysis procedure is developed to set the parameters for RLR hazard prediction. Online-wise, a two-dimensional normal model is developed to predict the vehicle's stop-go maneuver based on speeds at advanced detectors. Additionally, unlike most prediction models which are designed to minimize mean errors, our model identifies two types of errors, namely the false alarm and missed report. The capability of distinguishing these two types of errors is crucial to the effectiveness of dynamic systems. To quantify the trade-off between these two types of errors in the system design, a system operating characteristics (SOC) function is then defined. Performance of the proposed model and its prediction algorithm is evaluated using data collected from a field intersection. At a false alarm rate of 5%, the algorithm reach a correct detection rate of over 65% to over 90% for various legs of the test intersection. Performance evaluation results showed that the proposed models and algorithms within the DARE framework can effectively detect the RLR hazards.

Keywords: Traffic Control, Collision Avoidance System, Red-light running, Dynamic all-red extension, Probabilistic Model of Driver Behaviors

Executive Summary

Red light running has been a major cause of intersection injuries and fatalities in the United States. In 2004 alone, there were 8,619 fatal crashes and 848,000 crashes with people injured, all caused by RLR. Under the U.S. Department of Transportation (DOT) initiated Collaborative Intersection Collision Avoidance System CICAS program (Signalized Left Turn Assistance −SLTA and Traffic Signal Adaptation − TSA), California PATH program at the University of California, Berkeley undertake the efforts of designing a system that pro-actively sense the impending hazards that are caused by RLR and enact dynamic control of the signal by extending the all-red clearance interval by a few seconds to avoid potential hazardous situation.

In 2007, California PATH setup an observation field intersection at Page Mill road and El Camino Real at Palo Alto, with 9 Autoscope ® cameras instrumented to collect vehicle movement data as well as the real-time signal phase. The data collected in 2008 and 2009 have been used to study the statistical characteristics of the vehicle trajectories approaching the intersection during the inter-green phase (including yellow and red). Particularly we have used the collected data to analyze the following types of vehicle trajectories of (1) Red light runners, (2) First-to-stop vehicles, and (3) vehicles going through the intersection during yellow, to identify the statistically significant contributory parameters from the trajectories that can be used for RLR related hazard prediction purpose.

The data analysis results showed that important contributory factors that need to be included for accurate RLR hazard prediction should include the single vehicle trajectory parameters such as speed, time to intersection and acceleration / deceleration, and inter-vehicle status as whether or not the subject vehicle is following a leading vehicle. A probabilistic model that uses both the single vehicle parameters and inter-vehicle status has been built for the purpose of RLR hazard prediction.

Besides the probabilistic prediction model for RLR hazards, we have also developed a framework to optimize the parameter settings for the system using a so-called System Operating Characteristic (SOC) curve, which shows the trade off of two types of errors that could occur to the system: miss detection and false reports. We used a Neymann-Pearson criteria to gain the optimized parameters that achieve best performance under a constraint of maximum allowable false alarm rate.

Performance of the proposed system has been evaluated using the collected field data. The achieved performance (in terms of the correct detection rate of RLR hazards at different given levels of false report rates) has been evaluated.

Results of this research laid the basis for a operational dynamic all-red extension system, by having the following accomplishments:

- Built a set of contributory factors from the vehicle trajectories that can be used to predict RLR hazards;
- Built probabilistic models of the RLR hazards;
- Developed RLR hazard prediction method with optimized performance; and

- Evaluated actual performance of the system using field collected data.

The research results showed that the dynamic all-red extension as designed using the proposed framework, can predict the RLR hazards with high accuracy while also with a low false report rate at the same time. This dynamic all-red extension framework is ready for a field operational test that will address implementation issues related to deployment.

This report is for Caltrans Task order 65A0363 (Collaborative Intersection Collision Avoidance System).

Contents

List of Figures

List of Tables

1 Introduction

1.1 Red-Light Running

Red-light-running (RLR) is an increasingly major national safety issue at signalized intersections. In 2004, 2.5 million or 40% of all police-reported crashes in the United States occurred at or near intersections [14]. Of these intersection-related crashes 8,619 or 22.5% were fatal and 848,000 or 46% resulted in injury. The RLR aspect of this problem is known and serious [1, 5, 14], as approximately 20% of all intersection crashes occur due to signal violation crashes [5], resulting in an estimated monetization of $13 billion a year when considering fatalities, lost wages, medical costs, property damages, and insurance [7].

To address the RLR problem, significant effort has been undertaken to conceive engineering countermeasures. These findings are summarized by Retting et al., and they run the gamut from optimization of signal timing through removing unwarranted signals [17].

One important approach is to extend the yellow interval. It is reported that 1 second more yellow interval reduces as much as 50% of RLR, especially for those intersections where yellow intervals were set lower than the Institute of Transportation Engineers (ITE) standard [4]. For high speed intersections, Green Extension System (GES) has been developed and research shows that it helps to effectively reduce RLR caused by dilemma zone [2].

Despite of these significant countermeasures, RLR remains to be an outstanding cause for intersection collisions, partly due to the fact that significant portion of the RLR cases feature willful, intentional running or inattentive platoon following. These RLR scenarios lead to over 50% of the RLR crashes [11], among which most are against Left Turn Across Path from Opposite Direction (LTAP-OD) vehicles (usually during the first few seconds after red onset) and Straight Cross Path (SCP) vehicles (last until late into red).

1.2 Dynamic All-Red Extension

There have been significant research and development in recent years in the application of technology to address the intersection collision problem [1, 6, 7, 13, 14, 22]. The central concept of a dynamic system therein is to actively inform or alert drivers of emerging, impending hazardous situations.

Extending the all-red interval is an intuitive approach to counteract the RLR related crashes, especially those caused by willful runners or inattentive runners who enter the intersection during the first few seconds of red onset. A study carried out by researchers at Iowa State University shows that extending all-red interval does reduce the crash rate, however the drivers could potentially adapt to the extension and undo the safety improvement [19].

Through applying the aforementioned “dynamic” concept to RLR violations, a provocative solution, which potentially addresses the dilemma zone, willful violation and inattentive violation motivations is to “dynamically” extend an ”all red” phase. This therefore relies in large part on detection of potential hazardous situation caused by a signal violator to enact a Dynamic All-Red Extension (DARE) - and it is precisely this probabilistic detection and implementation into an algorithm on which we focus this report. Research reported in this report is for the U.S. De-

partment of Transportation (DOT) initiated Collaborative Intersection Collision Avoidance System CICAS program (Signalized Left Turn Assistance −SLTA and Traffic Signal Adaptation − TSA) conducted under the auspices of the University of California, Berkeley effort.

To elaborate: the detection (prediction) of RLR hazardous situation is a crucial part for the collision avoidance systems, which must be able to sense the dynamic characteristics of a vehicle approaching the intersection from a reasonable range. A dynamic "state-map" of the intersection defined by Zennaro and Misener can then be built upon the sensed components [21]. Instead of collision, which is so rare in both space and time such that its prediction is very difficult, we propose to predict hazardous situation based on the state map to enable the engineering measures to protect vehicles affected by the violator [2].

We employ a predictive encroachment time to quantify the RLR hazardous situation when DARE will be triggered if this time lapse is below a threshold, which for a given collision course can be transformed into a time-into-red threshold for RLR. Thus we address the problem of RLR hazard prediction with a combination of finding the threshold of time into red for a given collision course, predicting the vehicle stop-and-go motion plus estimating of the arrival time (time into red). The first part, finding the threshold of time into red for RLR hazard is based on data analysis of the empirical distribution of RLR and the entry time of cross street traffic. The second part, predicting the stop-go maneuver of vehicles, is a critical part of the system which involves predicting individual diver behavior at intersections. Sheffi and Mahmassani have introduced a statistical model wherein the driver decision to stop is normally distributed and a function of "critical time" [18]. More recent research in driver behavior during the intersection approach has used logit models [3, 8] or empirical models [15]. The results *in toto* show that the stopping probability is a function of multiple factors, to include the distance-to-intersection, speed and headway [3, 10, 18]. We extend Sheffi and Mahmassani's probabilistic model to multi-dimensional normal distribution where measured vehicle-to-intersection factors can be incorporated.

In short, we view the intersection hazard as an evolving event for various distances of the violator to intersection; therefore, the empirical distribution of entry time of the conflict traffic, the distance to intersection of violator, speeds of the approaching vehicles and time into yellow (red) are important contributory factors to our model.

When applying a prediction algorithm to the collision avoidance system, there are usually two types of errors: the missing report error when a RLR hazard is missed (false negative), and also the false alarm error when proceeding through before a preset critical time or stopping is accepted as RLR hazard (false positive). Clearly, in any RLR countermeasure there could be different consequences for the two types of errors, with the missing report giving a false sense of security and the false alarm being a nuisance; therefore, a means to minimize and control the balance is essential for any implementation. As an important note in implementing such a balance, we consider the *a priori* probability of hypotheses can be different. For example, at most intersections, the occurrence of RLR hazards could be significantly less than that of non-violation. Thus the requirement of false alarm rate should be at or lower than a given threshold.

This is in contrast with the traditional probabilistic analysis for RLR frequency estimation or dilemma zone prediction, as these do not distinguish the two types of errors. We address our methodology - or method to control the balance -with the Neyman-Pearson (NP) principle, which

governs the prediction of RLR hazard occurrence based on the empirical observations by maximizing the probability of detection (minimizing the missed detection rate) while keeping the false alarm rate equal to or lower than a given level [12]. The capability of adjusting the preset level of false alarm is important to the effective operation of a dynamic system.

We have restricted the system to use only the real-time point detection sensors. The reason for restricting the type of sensors is for deployability and cost considerations. Certainly, using or extending legacy systems makes the conceived system more viable. Therefore, our approach is to modify existing point detectors to serve as sensing means for RLR hazard prediction or by using emerging but currently-marketed products. As a case in point in this study, the emulated speed loops of Autoscope®cameras are used. They are virtual point detectors which give good real-time speed measurements. At the same time, the locations of the point detectors are software-reconfigurable, which makes it possible to optimize the location of the detectors based on empirical observations.

2 Dynamic All-Red Extension Outline

2.1 Predictive Indicator of RLR Hazard

The system is designed to be "dynamic", which should adaptively predict the occurrence of RLR and its related hazardous situation based on advance detection, and thereafter enacts all-red extension before all-red interval terminates, when most likely the subject violator has not yet entered the intersections. We need to develop a predictive indicator to address the severity of the potential hazardous situation caused by the signal violator. When this indicator shows that likelihood of collision to happen is high, the DARE will be triggered. The essential part of the research is to build a probabilistic model and upon which generate the threshold(s) for triggering the DARE.

Being aware of the rareness of the collision occurrences both in time and space, we adopt an alternative proximal indicator based on Encroachment Time (ET). ET is an objective indicator which is used extensively to quantify the severity of near-miss situations, and which, when below a certain level, indicates a high likelihood of crashing, the frequency of which is much higher than that of real crashes. Using conflicts instead of collisions is considered to be a reasonable alternative evaluation tool [20] [9].

Traditional ET and other variations of indicators such as Post-Encroachment Time (PET) defines scenarios when subject vehicle (SV) and principal other vehicle (POV) are both moving (at known speeds). While for DARE system, it needs to decide whether or not to extend the all-red interval before green onset for the conflict traffic when POV is mostly likely still waiting to enter the intersection with zero speed.

To address this difference, we define a derivative of ET / PET as:

the predicted time lapse between the SV clears the conflict zone and the a priori entry time of the POV to the conflict zone.

We call this time lapse the "Predicted Encroachment Time for RLR", or $\mathrm{PrET_{RLR}}$. $\mathrm{PrET_{RLR}}$ itself is a random variable which is a function of the speed of the SV, distance to conflict zone and the *a priori* probability distribution of entry time of the POV.

The system extends all-red to avoid collision when

$$\Pr(\text{PrET}_{\text{RLR}} < \text{threshold}) < \text{Minimum allowed probability of hazard}, \tag{1}$$

where "threshold" is a pre-set constant below which the conflict will be considered as hazardous situation likely to turn into collision.

Assume that violator clears the intersection in constant speed, then the condition defined in (1) is equivalent to a condition of

$$\text{Violator's time into red} < \tau, \tag{2}$$

where τ is a function of the speed of SV, distance to conflict zone and the *a priori* distribution of entry time of POV.

2.2 System Outline

The algorithm of the DARES system, is illustrated in Figure 1. It includes two parts, one is an offline parameter setting procedure which collects empirical data to get the parameters of the probabilistic models in (1) as well as in RLR prediction which will be stated in detail in next section. The second part is the online algorithm which predicts hazardous situations using advanced point detector outputs and the parameters extracted from the offline procedure.

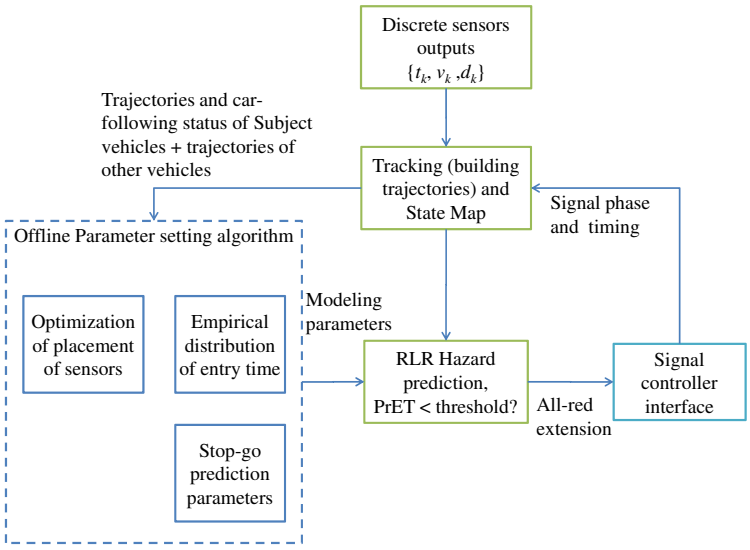


Figure 1: Dynamic All Red Extension Algorithm

3 Modeling and Neyman-Pearson Detection of RLR Hazard

3.1 Modeling of RLR Hazard

To build the probabilistic model of RLR hazard, we hereby define several quantities that are closely related to RLR and its hazards.

Definition 1 *RLR hazard.*

A RLR hazard is an event when, according to empirical statistics, the probability of $\text{PrET}_{\text{RLR}} < \delta_0$ (δ_0 is a given threshold) is greater than some allowed minimum value $P_{\min}$. A commonly used threshold for ET / PET in safety assessment is $\delta_0 = 1$ second.

Definition 2 *Clear time of the violator t_c and entry time of the conflict vehicle t_e.*

To avoid the needs to define details of vehicle trajectories for different intersection geometry, we instead define clear time t_c and entry time t_e which can be directly measured given the collision course (and thus the conflict zone). t_c is the time for the violator to leave the conflict zone and t_e is the time when the POV to enter the conflict zone. We define the time into red of the violator to be τ, then we have that (t_c and t_e are referenced to red onset)

$$\tau = t_c - t_{rc}, \tag{3}$$

where t_{rc} is the time duration for the violator to travel from the stop bar to clear the conflict zone. t_e has an empirical probability distribution ("Pr" stands for the probability of an event)

$$F_{t_e}(x) = \Pr\{t_e < x\}, \tag{4}$$

which can be estimated using histogram of historical data.

The condition in (1), given t_c known from the speed and distance to conflict zone of the violator, can be expressed that

$$\begin{aligned} &\Pr\{\text{PrET}_{\text{RLR}} < \delta_0 | t_c\} \\ =&\Pr\{t_e - t_c < \delta_0 | t_c\} \\ =&\Pr\{t_e < t_c + \delta_0 | t_c\} \\ =&F_{t_e}(t_c + \delta_0). \end{aligned} \tag{5}$$

The above probability being less than $P_{\min}$ is equivalent to

$$\begin{aligned} & F_{t_e}(t_c + \delta_0) \le P_{\min} \\ \rightleftharpoons \quad & \tau + t_{rc} + \delta_0 \le F_{t_e}^{-1}(P_{\min}) \\ \rightleftharpoons \quad & \tau \le F_{t_e}^{-1}(P_{\min}) - t_{rc} - \delta_0. \end{aligned} \tag{6}$$

To simplify the evaluation of time-into-red threshold in (6), we use the average speed of RLR for a given approach to get t_{rc},

$$t_{rc} = \frac{d_c}{\bar{v}_{\text{RLR}}}, \tag{7}$$

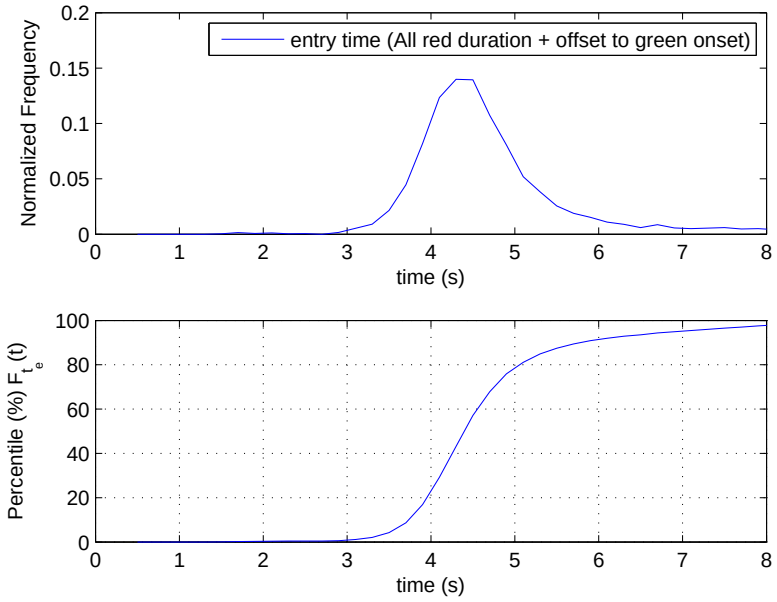


Figure 2: Example of entry time distribution

where d_c is the distance for the violator to clear the conflict zone from the stop bar and $\bar{v}_{\text{RLR}}$ is the average speed of RLR.

In Figure 3, the time space diagram of a conflict zone for which the calculation of the predictive encroachment time using the probabilistic distribution of the entry time and a trajectory of a RLR is illustrated.

Based on (6), we quantify the RLR hazard using the time into red of the violator. Therefore the essential part of the algorithm will be the prediction of RLR, which is in turn a combination of the prediction of the vehicle's stop-go maneuver and the estimation of its arrival time at the conflict zone.

For RLR hazard prediction, we are only interested in two types of maneuvers : the "first-to-stop" and "go through" (actually not all the going through vehicles, but those pass the point sensors during a certain period, say from yellow onset to the first few seconds after red onset). The first-to-stop vehicles are the ones at each lane that stopped first after the yellow onset. This maneuver are not driver-based taxonomy. They are coupled with observation of only the vehicles, therefore the need to define more microscopically driver-related information is obviated.

For each possible collision course, we place point detectors along the track of going through vehicles to collect data for RLR hazard prediction and along the track of the conflict vehicles (both SCP and LTAP-OD) to collect data for empirical distribution of entry time. In the point detectors along the going through track, we include one at the stop bar and at least one to monitor the clearance of conflict zone. While for modeling and prediction purposes, we picked up M out of these (for the model developed in this paper, $M = 2$), at distances

$$\{d_{k_1} \text{ and } d_{k_2}, d_{k_1} > d_{k_2}\}$$

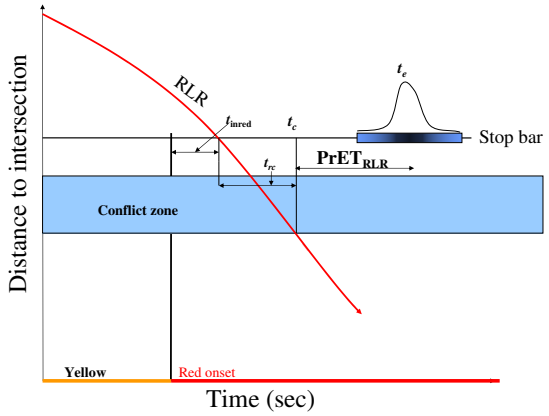


Figure 3: Time Space Diagram of a Conflict Zone and the Calculation of its PrET

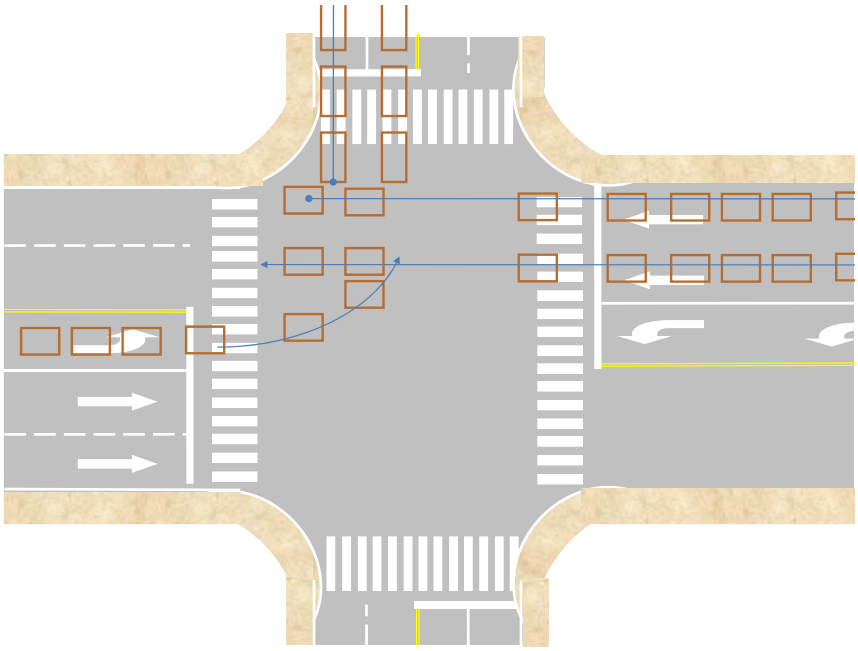


Figure 4: Discrete point sensors

from the intersection. The detected speeds at these locations are

$$\{v(k_1), v(k_2)\}, \tag{8}$$

and the time instant (relative to red onset) when these measurements are obtained

$$\{t(k_1), t(k_2)\}. \tag{9}$$

The time stamps in (9) are not independent, assuming that vehicles move at constant acceleration between the detectors. In this case, the data are highly dependent. Thus only the latest data point $t(k_2)$ will be used.

The arrival time at the stop bar estimated from the last detector is then

$$\hat{t}_I(k_2) = t(k_2) + \hat{t}_{2I} \tag{10}$$

where $\hat{t}_{2I}$ is the estimated travel time from current location to intersection.

First, we present a probabilistic model of the detection problem for the two types of scenarios.

The observations we have are

$$\{v(k_1), v(k_2), \hat{t}_I(k_2)\}. \tag{11}$$

We have two groups of hypotheses for the go-through (g) and first-to-stop (s). Among the go-through vehicles, the RLR and go-through-in-yellow vehicles only differ in their arrival time at the intersection, or by definition:

$$\begin{aligned} t_I > 0, &\quad \text{RLR}, \\ t_I \leq 0, &\quad \text{through in Yellow}. \end{aligned} \tag{12}$$

We define RLR hazard as arriving at an intersection at a time τ seconds after red onset, where τ is obtained using (3), thus

$$\begin{aligned} t_I > \tau, &\quad \text{RLR hazard}, \\ t_I \leq \tau, &\quad \text{non-hazard}. \end{aligned} \tag{13}$$

To further compact the variables used in the model, we define

$$\begin{aligned} \bar{v} &= \frac{1}{2}(v(k_1) + v(k_2)), \\ a(k_2) &= \frac{v(k_2) - v(k_1)}{t(k_2) - t(k_1)}, \end{aligned} \tag{14}$$

the two-dimensional observation

$$\mathbf{X} \triangleq \{x_1 = a(k_2), x_2 = \bar{v}\}^T, \tag{15}$$

and the clear time t_c.

Because many random factors contribute, $\mathbf{X}$ can be reasonably hypothesized to be a bi-variate normal random vector. Thus to formulate the hypotheses, we have

$$\begin{aligned} \mathbf{H_s}: \ \mathbf{X} &\sim \mathcal{N}(\mathcal{U}_\mathbf{s}, \Sigma_\mathbf{s}) \\ \mathbf{H_g}: \ \mathbf{X} &\sim \mathcal{N}(\mathcal{U}_\mathbf{g}, \Sigma_\mathbf{g}) \end{aligned} \tag{16}$$

where $\mathbf{H_s}$ and $\mathbf{H_g}$ are hypotheses for "stopping" and "going", $\mathcal{U}_\mathbf{s}$, $\Sigma_\mathbf{s}$, $\mathcal{U}_\mathbf{g}$, $\Sigma_\mathbf{g}$ are the mean and covariance matrices of $\mathbf{X}$ under the two hypotheses, respectively.

3.2 Predictor of RLR Hazard and Two Types of Errors

To derive the prediction method for RLR, we generate several definitions regarding RLR errors.

Definition 3 *Predictor of vehicle maneuvers.*

A detector which predicts the vehicle maneuvers gives a binary decision between the hypotheses "stop" or "go", or

$$\mathbf{D}(\mathbf{X}) = \left\{\mathbf{H_s} \text{ or } \mathbf{H_g}\right\}. \tag{17}$$

which distinguishes two hypotheses: $\mathbf{H_s}$ and $\mathbf{H_g}$ using $\mathbf{X}$ which includes the average speed and speed difference (acceleration) of the violator. The detection would present two types of errors, the probabilities of which are respectively

$$\text{false alarm:} \quad P_{\text{FA}}(g|s) \triangleq P\left\{\mathbf{D}(\mathbf{X}) = \mathbf{H_g}|\mathbf{H_s}\right\}, \tag{18}$$

$$\text{and missing:} \quad P_{\text{M}}(s|g) \triangleq P\left\{\mathbf{D}(\mathbf{X}) = \mathbf{H_s}|\mathbf{H_g}\right\}. \tag{19}$$

Definition 4 *Time to stop bar estimator.*

We need to estimate the arrival time at the stop bar when the vehicle passes through the last discrete sensor at $d(k_2)$. To minimize the chance of missing any RLR hazard due to an improperly estimated arrival time, we need to construct an estimator $\hat{t}_I(k_2)$ which satisfies the following condition that

$$\Pr\left\{\hat{t}_I(k_2) \leq \tau | t_I > \tau, \mathbf{H_g}\right\} \simeq 0. \tag{20}$$

Definition 5 *Detector (predictor) of RLR Hazard.*

We denote the detector of RLR hazard as $\mathbf{D}_{\text{H}}(\mathbf{X})$. It can be defined as

$$\mathbf{D}_{\text{H}}(\mathbf{X}) = \begin{cases} 1, & \text{if } \mathbf{D}(\mathbf{X}) = \mathbf{H_g} \text{ and } \hat{t}_I(k_2) > \tau, \\ 0, & \text{otherwise.} \end{cases} \tag{21}$$

Based on the results of stop-go detector $\mathbf{D}(\mathbf{X})$, there are two possible situations when $\mathbf{D}_{\text{H}}$ misses a hazard, when a hazard is missed by $\mathbf{D}(\mathbf{X})$ or when, although $\mathbf{D}(\mathbf{X})$ reports correctly, the arrival time estimator misses the hazard. Thus, we can formulate the missing probability as

$$\begin{aligned} P_{\text{M-H}} \triangleq & \Pr\left\{\mathbf{D}(\mathbf{X}) = \mathbf{H_s}|\mathbf{H_g}, t_I > \tau\right\} \\ & + \Pr\left\{\mathbf{D}(\mathbf{X}) = \mathbf{H_g}, \hat{t}_I(k_M) \leq \tau | \mathbf{H_g}, t_I > \tau\right\}. \end{aligned} \tag{22}$$

The false alarm of RLR hazard happens either when the arrival time predictor falsely indicates a through vehicle before τ seconds into red as a hazard or when the $\mathbf{D}(\mathbf{X})$ falsely treats a stopping vehicle as not to stop.

$$\begin{aligned} P_{\text{FA-H}} \triangleq & \Pr\left\{\mathbf{D}(\mathbf{X}) = \mathbf{H_g}, \hat{t}_I(k_M) > \tau, t_I \leq \tau, \mathbf{H_g}\right\} \\ & + \Pr\left\{\mathbf{D}(\mathbf{X}) = \mathbf{H_g}, \hat{t}_I(k_M) > \tau, \mathbf{H_s}\right\}. \end{aligned} \tag{23}$$

Table 1: Numerical examples of missing rate and false alarm rate

	Samples	Hazard	Missed	$P_{\text{M-H}}$	false reports	$P_{\text{FA-H}}$
$\mathbf{H_g}$	600	10	1	10%	5	1%
$\mathbf{H_s}$	400	/	/		5	

It is a noteworthy fact that definitions of probabilities of missing and false alarm in (22) and (23) are different, the first one being defined using conditional probability while the other one using un-conditional probability. A numerical example should explains why. For the example in Table 1, we missed one hazard out of 10 total hazards and 1000 total samples of interest, in which case the missing rate should be 10% which is the ratio conditioned on the number of hazards only. While for false alarm rate, since all the samples (except for hazards) could be dedicated for it, it is calculated over the whole sample size then.

3.3 Neyman-Pearson Detection of RLR Hazard

It is our wish to quantitatively evaluate the DARE system in terms of the error performance. Neyman-Pearson criterion is well fitted into this need. It says that we should construct the decision rule to have maximum probability of correct detection while not allowing the probability of false alarm to exceed a certain value. In other words, we want to have a system performs in the following way

$$\begin{aligned} &\max(1-P_{\text{M-H}}), \\ &\text{such that } P_{\text{FA-H}} \leq \theta, \end{aligned} \tag{24}$$

where θ is a preset threshold of allowable false alarm rate.

Then the problem turns into *how can we evaluate the error probabilities in (22) and (23)?* The direct evaluation of them, however, are difficult due to their multi-variate nature. To simplify the problem, we use approximate expressions to the error probabilities and link them to $P_{\text{M}}(g|s)$ and $P_{\text{FA}}(s|g)$, which under the two-dimensional normal model are easier to evaluate.

3.3.1 Approximation of $P_{\text{M-H}}$

There are two terms in its definition (22). The second part, while with the assumption (20) that our arrival time estimator should avoid such missing to happen, can be neglected. Also since our stop-go model depends only on the speed ($\bar{v}$) and acceleration ($a(k_2)$), it can be reasonably assumed that $\mathbf{D}(\mathbf{X}) = \mathbf{H_s}$ is independently distributed to the arrival time for our interested samples (those arrive late in yellow or early in red). Therefore, we have that

$$\begin{aligned} P_{\text{M-H}} \simeq &\Pr\left\{\mathbf{D}(\mathbf{X}) = \mathbf{H_s} | \mathbf{H_g}, t_I > \tau\right\} \\ \simeq &\Pr\left\{\mathbf{D}(\mathbf{X}) = \mathbf{H_s} | \mathbf{H_g}\right\} \\ = &P_{\text{M}}(s|g). \end{aligned} \tag{25}$$

3.3.2 Approximation of $P_{\text{FA-H}}$

$$P_{\text{FA-H}} = \underbrace{\Pr\left\{\mathbf{D}(\mathbf{X}) = \mathbf{H_g}, \hat{t}_I(k_2) > \tau, t_I \le \tau, \mathbf{H_g}\right\}}_{(1)} + \underbrace{\Pr\left\{\mathbf{D}(\mathbf{X}) = \mathbf{H_g}, \hat{t}_I(k_2) > \tau, \mathbf{H_s}\right\}}_{(2)},$$

where

$$\begin{aligned}(1) \simeq & \Pr\left\{\mathbf{D}(\mathbf{X}) = \mathbf{H_g}, \mathbf{H_g}\right\} \cdot \Pr\left\{\hat{t}_I(k_2) > \tau, t_I \le \tau\right\} \\ = & \Pr\left\{\mathbf{D}(\mathbf{X}) = \mathbf{H_g} | \mathbf{H_g}\right\} \cdot \Pr\left\{\mathbf{H_g}\right\} \\ & \cdot \Pr\left\{\hat{t}_I(k_2) > \tau, t_I \le \tau\right\} \\ = & (1 - P_{\text{M}}(s|g)) \cdot \Pr\left\{\mathbf{H_g}\right\} \\ & \cdot \Pr\left\{\hat{t}_I(k_2) > \tau, t_I \le \tau\right\}.\end{aligned}$$

approximately we have that

$$(1) \simeq (1 - P_{\text{M}}(s|g)) \cdot \Pr\left\{\mathbf{H_g}\right\} \cdot P_c,$$

where

$$P_c \triangleq \Pr\left\{\hat{t}_I(k_2) > \tau, t_I \le \tau\right\}. \tag{26}$$

We also have that

$$(2) \simeq P_{\text{FA}}(s|g) \cdot P(\mathbf{H_s}).$$

Now we have the simplified form

$$P_{\text{FA-H}} \simeq P_{\text{FA}}(g|s) \cdot P(\mathbf{H_s}) + (1 - P_{\text{M}}(s|g)) \cdot \Pr\left\{\mathbf{H_g}\right\} \cdot P_c. \tag{27}$$

Now that the performance of the DARE system is linked to the performance of its core algorithm: the stop-go maneuver predictor via the *a priori* probabilities of P_c, $\Pr\{\mathbf{H_g}\}$ and $\Pr\{\mathbf{H_s}\}$ which can all be estimated from empirical data. The following diagram shows how the DARE is optimized using the Neyman-Pearson criterion.

$$\begin{array}{lll} \text{Hazard Detection:} & P_{\text{FA-H}} = \theta & \xleftarrow{\theta = \beta \cdot P(\mathbf{H_s}) + (1 - P_{\text{M}}(s|g)) \cdot \Pr\{\mathbf{H_g}\} \cdot P_c} \quad \min P_{\text{M-H}} \\ & & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \uparrow P_{\text{M-H}} \simeq P_M(s|g) \\ \text{Stop-go Detection:} & P_{\text{FA}}(s|g) = \beta & \xrightarrow{\text{minimizing} P_M(s|g)} \qquad \min P_M(s|g) \end{array}$$

4 The Implementation of the RLR Hazard Prediction Algorithm

As has been illustrated in Figure 1, the DARE algorithm consists of two parts: the online prediction and the offline parameter setting.

- *Offline parameter setting:* We need to collect (huge amount of) data to get the parameters in the model. These parameters include the threshold of time-into-red of the violator, τ for each collision course, the *a priori* probabilities of Stop-Go maneuvers during the observation window ($P(\mathbf{H_g}$, $P(\mathbf{H_s})$) and also P_c. Then offline optimization should be carried out the obtain the boundary of stop-go maneuver prediction (v_0, a_0).
- *Online RLR hazard prediction:* Online procedure is simple and straightforward. All-red extension is triggered when a vehicle is predicted to run red light with time into red greater than given threshold for the most likely collision course.

4.1 Offline procedure of DARE algorithm

4.1.1 Estimating the parameters for prediction

When estimating the algorithm parameters, not only the sensor data at $d(k_1), d(k_2)$ are used (or $\mathbf{X}$), but also the whole trajectories from the vehicle approaching the entering, entering and exiting the intersection are used. Note the whole set of sensor outputs of an interesting trajectory as Z_n, where X_n is the $\mathbf{X}$ vector (speed and acceleration) extracted from Z_n. We also need trajectories of conflict traffic $\{Y_m\}$.

Assume the size of set $\{X_n|\mathbf{H_g}\}$ is N_g and $\{X_n|\mathbf{H_s}\}$ is N_s, respectively. The mean and variance matrices in Algorithm 1 can be estimated, using

$$\hat{\mathcal{U}}_g = \frac{1}{N_g}\sum_n (\mathbf{X}_n|\mathbf{H_g}), \tag{28}$$

and

$$\hat{\blacksquare}_g = \frac{1}{N_g - 1}\sum_n (\mathbf{X}_n|\mathbf{H_g})(\mathbf{X}_n^T|\mathbf{H_g}). \tag{29}$$

For $\mathbf{X}_n|\mathbf{H_s}$, the estimator is similar.

4.1.2 Forming the Neyman-Pearson Detection Boundary

We define $P\{X|\mathbf{H_g}\}$ and $P\{X|\mathbf{H_g}\}$ as the probability density functions (PDF) of $\mathbf{X}$ under hypotheses $\mathbf{H_s}$ and $\mathbf{H_g}$, respectively (see (16)), β as the false alarm rate for $\mathbf{D}(\mathbf{X})$. To form the Neyman-Pearson detector which minimizes the missing probability while keeping the false alarm rate not greater than β, the following constrained optimization problem need to be solved:

$$\begin{aligned} &\max \left(\int_{\mathbf{X}\in\mathbf{R_f}} P\{\mathbf{X}|\mathbf{H_g}\} d\mathbf{X} \right), \\ &\text{given any } \mathbf{R_f}, \text{ such that } \int_{\mathbf{X}\in\mathbf{R_f}} P\{\mathbf{X}|\mathbf{H_s}\}\, d\mathbf{X} < \beta. \end{aligned} \tag{30}$$

Algorithm 1: Estimating parameters for prediction

input: A set of interesting trajectories $\{Z_n, 1 \leq n \leq N\}$ of size N ;
A set of trajectories $\{Y_m, 1 \leq m \leq M\}$
from conflict path ;
foreach Z_n **do**
  if Z_n *is RLR* **then**
    $X_n \in$ RLR ;
  end
end
Calculate the average speed of RLR $\bar{v}_{\text{RLR}}$;
Get the histogram of entry time t_e using all first-to-enter in green from Y_n ;
Get $F_{t_e}(x)$ from the histogram ;
For SCP and LTAP-OD, use (6) to calculate τ for each conflict course and choose the minimum one corresponding to each scenario ;
foreach Z_n **do**
  if Z_n *is going through* **then**
    $X_n \in \mathbf{H_g}$;
    calculate $\hat{t}_I$ using X_n ;
    if $\hat{t}_I > \tau$ *and* $t_I < \tau$ **then**
      $X_n \in \{\hat{t}_I > \tau, t_I < \tau\}$;
    end
  else
    X_n is first-to-stop ;
  end
end
We have the sets $\{X_n|\mathbf{H_g}\}, \{X_n|\mathbf{H_s}\}$;
Estimate $P(\mathbf{H_g})$ and $P(\mathbf{H_s})$ using their frequency respectively ;
Estimate P_c using the frequency of $X_n \in \{\hat{t}_I > \tau, t_I < \tau\}$;
Estimate the mean and covariance matrices for hypothesis $\mathbf{H_g}$ and $\mathbf{H_s}$;
Result: τ, $\hat{P}(\mathbf{H_g})$, $\hat{P}(\mathbf{H_s})$, $\hat{P}_c$ and $\mathcal{N}(\hat{\mathcal{U}}_g, \hat{\Sigma}_g)$, $\mathcal{N}(\hat{\mathcal{U}}_s, \hat{\Sigma}_s)$

Given a false alarm rate β, we need to search the observation space to find the solution to (30). For the multi-variate detection problem here, there exist various kinds of reasonable decision boundaries, which usually requires a computationally prohibitive search to find the optimal solution. For simplicity and clear physical interpretation, we adopt the rectangular boundary

$$\begin{aligned} \bar{v} &> v_0, \\ a(k_2) &> a_0, \end{aligned} \tag{31}$$

when we find a sub-optimal solution to the multi-variate optimization problem.

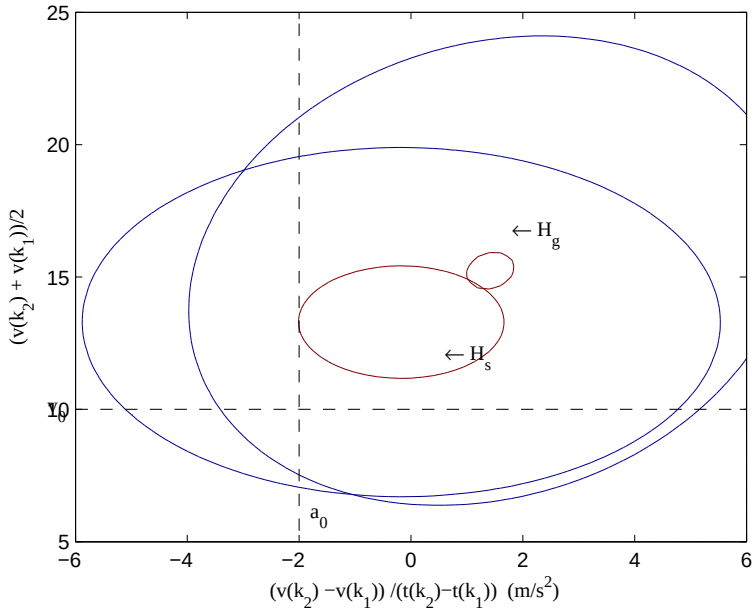


Figure 5: Bi-variate Normal model of the two hypotheses

The Neyman-Pearson detection (30) can be formed by finding the boundary variable $\{v_0, a_0\}$ according to the given false alarm rate θ. That is, to solve the following problem:

$$\{v_0, a_0\} = \arg \max_{\substack{\bar{v}_{\min}<\bar{v}<\bar{v}_{\max} \\ a_{\min}<a(k_2)<a_{\max}}} \int_{\substack{\bar{v}>x_1 \\ a(k_2)>x_2}} \Pr\{\mathbf{X}|\mathbf{H_g}\}\, d\mathbf{X}, \tag{32}$$

where

$$\text{for all } x_1 \in (\bar{v}_{\min}, \bar{v}_{\max}),\, x_2 \in (a_{\min}, a_{\max}) \text{ such that } \int_{\substack{\bar{v}>x_1 \\ a(k_2)>x_2}} \Pr\{\mathbf{X}|\mathbf{H_s}\}\, d\mathbf{X} \le \beta, \tag{33}$$

where β is the false alarm rate set according to θ such that

$$\theta = \beta \cdot P(\mathbf{H_s}) + (1 - P_{\mathrm{M}}(s|g)) \cdot \Pr\{\mathbf{H_g}\} \cdot P_c. \tag{34}$$

To sum up, we have the following algorithm.

Algorithm 2: Forming the Neyman-Pearson detection boundary

input: Given maximum false alarm rate θ_0 and the parameters obtained in Algorithm 1.
for $0 < \beta < 1$ **do**
  foreach *a in* $(a_{\min}, a_{max})$ *and* $\bar{v}$ *in* $(\bar{v}_{\min}, \bar{v}_{\max})$ **do**
    Calculate $P_{\mathrm{FA}}(g|s)$ and $P_{\mathrm{M}}(s|g)$ using estimated mean and variance ;
    if $P_{FA}(g|s) \leq \beta$ **then**
      find the $(a_0, \bar{v}_0)$ pair that minimizes $P_{\mathrm{M}}(s|g)$;
    end
  end
  Convert false alarm rate of hazard detection β to θ of stop-go predictor using (34);
  Find the β that makes $\theta(\beta) = \theta_0$;
end
Result: $(a_0, \bar{v}_0)$ that minimizes $P_{\text{M-H}}$ and also keeps $P_{\text{FA-H}} \leq \theta_0$.

4.2 Online Prediction of RLR Hazard

4.2.1 Arrival time (time-into-red) estimator

The arrival time estimator $\hat{t}_I(k_2)$ need to meet the assumption in (20). To incorporate the measured acceleration $a(k_2)$, we adopt an estimator in the following form

$$\hat{t}_I(k_2) \triangleq t(k_2) + \frac{d(k_2)}{v(k_2)} + \rho[a(k_2) - \bar{a}(k_2)], \tag{35}$$

where $\bar{a}(k_2)$ is the average of measured acceleration of the RLR samples collected, $\rho < 0$ is a constant coefficient which rejects some vehicles make to go-through legally by accelerating after yellow onset, while $|\rho|$ should be kept small and fine tuned to meet requirement (20). Empirical data show that $\rho = -0.05 s^3/m$ is a good option (a is in m/s^2 and t in seconds).

4.2.2 Formulation of $\mathbf{D_H}(\mathbf{X})$

Having obtained the optimized decision boundary and arrival time estimator, $\mathbf{D}_{\mathrm{H}}(\mathbf{X})$ can be formulated as

$$\begin{aligned} \mathbf{D}_{\mathrm{H}}(\mathbf{X}) &= 1, \text{ if } a(k_2) > a_0 \text{ and } \bar{v}(k_2) > v_0 \text{ and } \hat{t}_I(k_2) > \tau \\ &= 0, \text{ otherwise.} \end{aligned} \tag{36}$$

4.3 Performance of the RLR Predictor

The probability of detection (defined as $P_{\mathrm{D}} \triangleq 1 - P_{\mathrm{M}}$) as a function of probability of false alarm, is frequently used to quantify the performance of detection systems. With radar systems, where Neyman-Pearson criteria has been extensively used, this function is called the receiver operating characteristics (ROC). For an intersection collision avoidance system, we will define a similar

system operating characteristics (SOC) function as

$$\begin{aligned} \mathrm{SOC}(\theta) &\triangleq \max P_{\mathrm{D}}\left(\mathbf{D}_{\mathrm{H}}(\mathbf{X})\right), \\ &\text{where } P_{\text{FA-H}}\left(\mathbf{D}_{\mathrm{H}}(\mathbf{X})\right) \leq \theta. \end{aligned} \tag{37}$$

5 Data Collection and Processing

5.1 Field Setup and Configuration of Point Detectors

Field data are collected using the Autoscope cameras. Nine cameras are installed at three approaches of an arterial intersection (Page Mill Road and El Camino Real, Caltrans State Route 82) in the San Francisco Bay Are. We note that the signal phase and timing setting of this intersection fully conforms to the ITE standard. Three cameras are installed for each approach, covering respectively the exit area, stop area and advance area. We summarize the configurations into Table 2.

Table 2: Configuration of the field intersection

Page Mill rd. at El Camino Real			
	East bound	West bound	South bound
Lanes (recorded)	2	1	2
Speed Limit (mph)	30	35	35
Average Speed (mph)	30	35	35
90% speed less than (mph)	45	45	40
Yellow Duration (s)	4	4	4
All-red Interval (s)		0.5	
# Speed Loops (Exit)/ Lane	1	1	1
# Speed Loops (advance) / Lane	3	3	3
# Speed Loops (stop / conflict entry) / Lane	3	3	2
95% RLR time-into-red less than (s)	2.5s	2.5s	2.5s
Description of the approach	0.15 Mile from upstream intersection	No upstream intersection for More than 0.7 mile upstream	0.25 mile from upstream intersection, main street

As can be seen in Table 2, multiple virtual speed loops are placed at advance area of difference approaches to monitor the violation and also at stop area to get the conflict entry time behaviors.

5.1.1 PrET_{RLR} and τ

When there is a signal violator who is going through the intersection, there are possibly multiple conflict courses. For example, if the RLR is going South bound, first vehicle from each lane to East bound are all in SCP collision course. Since we are dealing with RLR happening in the first few seconds in red, the vehicle entering from outmost lane (furthest from runner) is more likely to be involved in a hazard.

For each approach, we obtain the empirical distribution of entry time $F_{t_e}(x)$ (see example of South Bound in Figure 2). τ is then calculated for each approach using the minimum PrET_{RLR} among different possible courses. Results are shown in Table 3.

Table 3: τ for each approach

Page Mill rd. at El Camino Real $-\ \tau$			
	East bound (LTAP-OD)	West bound (SCP)	South bound (SCP)
Conflict Approach	WB	SB	EB
d_c **(feet)**	96	110	120
entry distance to conflict (feet)	15	15	30
All red interval		0.5s	
$\textbf{PrET}_{\textbf{RLR}}$ **threshold**		1s	
$F_{t_e}^{-1}(P_{\min} = 0.3)$	5.2s	4.0s	4.2s
$\tau(P_{\min} = 0.3)$	1.5s	0.4s	0.5s

5.1.2 Minimum sensing distance requirement for DARE system

We can expect that the closer the distances of the virtual loops for prediction (namely $d(k_1)$ and $d(k_2)$) are, the higher the prediction accuracy will be. However, there is also a critical requirement that when a violator passes the last virtual loop, the system should be able to react (hold all-red), most of the time. This requires us to put the sensors far enough to cover most of the RLR, from which the *minimum distance requirement for DARE system* is formed as:

the last sensor should be far enough for over 95% of the violator (assumed to move at speed limit) to pass it before the termination of all-red interval.

There are always some RLR cases when the violator runs much late into red when there is no way to hold all-red for that long time. We try to cover most RLR during the first few seconds after red onset, to include the willful violator and inattentive platoon followers.

Results of the configurations of the virtual loops at the field intersection are summarized in Table 4. We note that "distance" indicates the distance of sensor from the stop bar to upstream direction (in feet). When less than zero, it indicates a sensor inside the intersection.

Table 4: Configurations of the Autoscope virtual loops

Page Mill rd. at El Camino Real : Sensor Configurations			
	East bound	West bound	South bound
Average Speed of RLR	30 mph	35 mph	38 mph
Min distance requirement	90 feet	110 ft	110 ft
$d(k_1)$ **and** $d(k_2)$	97,120	139,190	120,195
Locations of All Speed Loops	-110, -15, 0, 15, 40, 63, 97, 120, 190	-110, -15, 0, 20, 40, 95, 139, 190	-120, -30, 0, 30, 90, 120, 195

5.2 Data Processing

5.2.1 Calibration of the Sensors

The speed loops of Autoscope records the speed when a vehicle arrives at the specific location. To verify the accuracy of the engineering data generated from the Autoscope processors, we drove a fully-instrumented vehicle with GPS and data recorder at the test site for two days. The calibrated results are compared to the wheel speed, showing that the speed errors are usually less than 10%.

To validate the measurements from Autoscope, part of the RLR data are also compared to the recorded video clips. The comparison shows that the engineering data of Autoscope are consistent with the raw video data.

5.2.2 Building Trajectories From Point Detectors

To associate the data of discrete point sensors to a vehicle, multiple-hypotheses tracking (MHT) algorithm developed by Reed is employed [16]. MHT is a widely used tracking method which basically assumes that each discrete measurement could possibly be assigned to a previous found target (vehicle) with an *a priori* probability. Whenever a new measurement is reported by the sensor, it can be associated with (assigned to) any previous targets (vehicles), thus forming a set of new hypotheses. The MHT algorithm then calculates the evolving probability of each new hypothesis. Due to the exponential nature of this propagation, usually only the several most likely hypotheses are kept for further calculation, and finally (for example, when measurement from exit area has been associated to some track), hypothesis with highest probability is accepted.

6 Simulation

Computer simulation is carried out to verify the RLR hazard prediction algorithm. Data collected from the field are used to generate the parameters for each different approach of the site. Additional data are collected to verify the algorithm.

6.1 Simulation Parameters

Data are collected during August to September 2008. The parameters of the algorithm are obtained using Algorithm 1 and Algorithm 2.

Table 5: Prediction algorithm parameters

Page Mill rd. at El Camino Real			
RLR Hazard Occurrence	12	21	112
$\mathbf{H_s}$ Occurrences	912	462	2323
$\mathbf{H_g}$ Occurrences	4318	2725	11359
P_c	0.015	0.015	0.018
$\hat{P}(\mathbf{H_s})$	0.17	0.14	0.17
$\hat{\Sigma}_g$	$\begin{pmatrix} 4.7 & 1.6 \\ 1.6 & 10.2 \end{pmatrix}$	$\begin{pmatrix} 2.1 & 0.1 \\ 0.1 & 5.6 \end{pmatrix}$	$\begin{pmatrix} 4.8 & 1.4 \\ 1.4 & 13.0 \end{pmatrix}$
$\hat{\mathcal{U}}_g$	$\begin{pmatrix} 0.1 \\ 15 \end{pmatrix}$	$\begin{pmatrix} 0.5 \\ 13.2 \end{pmatrix}$	$\begin{pmatrix} 1.4 \\ 15.2 \end{pmatrix}$
$\hat{\Sigma}_s$	$\begin{pmatrix} 4.2 & -1.5 \\ -1.5 & 5.2 \end{pmatrix}$	$\begin{pmatrix} 4.1 & -1.6 \\ -1.6 & 5.3 \end{pmatrix}$	$\begin{pmatrix} 4.9 & -0.1 \\ -0.1 & 6.5 \end{pmatrix}$
$\hat{\mathcal{U}}_s$	$\begin{pmatrix} -1.9 \\ 13 \end{pmatrix}$	$\begin{pmatrix} -1.4 \\ 12.4 \end{pmatrix}$	$\begin{pmatrix} -1.1 \\ 13.3 \end{pmatrix}$

6.2 Simulated System Performance

Data collected during October 2008 are used to verify the algorithm. Simulation results using collected field data are compared to the expected results from the two-dimensional normal model.

6.2.1 Simulation procedure

RLR hazard prediction is simulated using parameters which are generated from Algorithm 1 and Algorithm 2. The data used to generate these model parameters are different than the data used for simulation, which is the intention of the simulation to also verify the consistency of the model with the time. The parameters setting procedure generated results are tabulated in Table 5. The RLR hazard parameters, including the τ for each approach is summarized in Table 3. The simulation procedure is described in Algorithm 3.

6.2.2 Simulation results

The actual SOC curves for the field intersection show that the simulated online performance of prediction algorithm (Figure 6), which is in good match with the theoretical SOC curves calculated using the model parameters. It is also noteworthy that the differences in the statistics for different travel directions lead to different system performance. This difference is most likely caused by variations in traffic with different approaches.

Algorithm 3: Simulation procedure

input: Model parameters listed in Table 5 and Table 3. Training data set (used to get the parameters) ;
New data set (other than the data used for parameter setting) ;
for $\theta = 0 \ldots 1$ **do**
  Use Algorithm 2 to Find the Neyman-Pearson Detection boundary $\{\bar{v}_0, a(k_2)_0\}$ with training data set ;
  Form $\mathbf{D}_{\mathrm{H}}(\mathbf{X})$ using $\{\bar{v}_0, a(k_2)_0\}$ and τ from Table 3 ;
  foreach *Data sample* X_n *from new data set* **do**
    if X_n *is hazard and* $\mathbf{D}_H(\mathbf{X}) = 0$ **then**
      This is a missing report;
    end
    if X_n *is not hazard and* $\mathbf{D}_H(\mathbf{X}) = 1$ **then**
      This is a false alarm ;
    end
  end
  Record the simulated false alarm rate and missing rate ;
  Use the two dimensional normal model to calculate (using numerical integration) the theoretical missing rate given θ ;
end
Result: Simulated SOC curve and the theoretical SOC curve.

6.2.3 Simulated system performance with different hazard settings

All the previous simulation results are obtained using one set of threshold for PrET_{RLR} (one second) and $P_{\min} = 0.3$. If these settings are changed, we need to go through the offline parameter setting procedure to get new parameters for the system. Either increasing the threshold $P_{\min}$ or decreasing δ_0 will increase the time-into-red threshold τ for RLR hazard, which will in turn reduce the rate when DARE triggering AND the false alarm rate as well (since the arrival time has less chance to make a mistake). Through simulation, we also verified this trend. In Figure 7, we plot the simulation results (simulated SOC curves) for South Bound with various time-into-red thresholds (τ), from 0.3s to 1.5s. When τ becomes bigger, the number of RLR hazard occurence (thus the rate that DARE system is triggered) is reduced drastically. And the improvement that the false alarm rate reduces with the increasing of τ is also observed.

7 Conclusion

Probabilistic model and prediction algorithm have been developed for a dynamic all-red extension system, to counteract the intersection safety hazard caused by RLR. We defined the PrET_{RLR} which is the time lapse between the time for the violator to clear the conflict zone and the *a priori* entry time of the first-to-enter vehicle from conflict approach. The hazard, which is determined by the intersection layout, empirical entry time of conflict traffic and speed of violator, is linked to an important characteristic of RLR: the time into red of the violator. We then developed a model using a minimum set of two point sensors to identify the RLR hazard. The prediction of the hazard is governed by the Neyman-Pearson criterion which gives the algorithm the capability of distinguishing two types of errors, namely the missing report and false alarm, and balance of their trade-off.

Field data collection is carried out to study the driver behavior related to RLR hazard and to set the parameters of the developed model / algorithm. The collected field data also serve the purpose of verifying the algorithm using simulation. System performance, measured by the trade-off of the two types of prediction errors, is calculated using field data and compared to the expected results of the model. A good match is found from this simulation which thereby verified the validity of the probabilistic model.

Performance of the proposed model and its prediction algorithm is evaluated using data collected from a field intersection. At a false alarm rate of 5%, the proposed algorithm reached correct detection rates of over 65% for West bound, 80% for South bound and over 90% for East bound, respectively, at the test intersection. The false alarm rate of 5% at the test intersection is equivalent to on average one false triggering per more than 3 hours. Performance evaluation results showed that the proposed DARE framework can effectively detect the RLR hazards.

There are also restrictions of employing DARE to address the RLR collision avoidance problem, mainly due to the concern that drivers may adapt to the system and the benefits of the system may be undone. We believe that via improving the hazard detection algorithm we could keep the triggering rate of DARE to a very low level when only really "dangerous" hazard will trigger the system thus minimize the influence on driver behaviors. Also the combination of DARE with enforcement such as Red-Light running Camera (RLC) could also be a good intersection safety

solution.

References

[1] J. Bonneson, D. Middleton, K. Zimmerman, H. Charara, and M. Abbas. Development and evaluation of a detection-control system for rural intersections. Federal Highway Administation FHWA/TX-02-4022-1, U.S. Department Transportation, Washington D.C., 2001.

[2] J. Bonneson, K. Zimmerman, and M. Brewer. Engineering countermeasures to reduce red-light running. Federal Highway Administration FHWA/TX-03/4027-2, Texas Transportation Institute, Texas, August 2007.

[3] J. A. Bonneson and H. J. Son. Prediction of expected red-light running frequency at urban intersections. *Transportation Research Record: Journal of the Transportation Research Board*, 1830:38–47, 2003.

[4] J. A. Bonneson and K. H. Zimmerman. Effect of yellow-interval timing on the frequency of red-light violations at urban intersections. *Transportation Research Record: Journal of the Transportation Research Board*, 1865:20–27, 2004.

[5] N. Brittany, B. N. Campbell, J. D. Smith, and W. G. Najm. Analysis of fatal crashes due to signal and stop sign violations. Report No DOT HS-809-779, National Highway Traffic Safety Administration, Washington D.C., 2004.

[6] C.-Y. Chan. Characterization of driving behaviors based on field observation of intersection left-turn across-path scenarios. *IEEE Trans. Intell. Transp. Syst.*, 7(3):322–331, Sept. 2006.

[7] J. Chang, D. Cohen, L. Blincoe, R. Subramanian, and L. Lombardo. CICAS-V research on comprehensive costs of intersection crashes. Report No 07-0016, National Highway Traffic Safety Administration, Washington D.C., 2007.

[8] T. J. Gates, D. A. Noyce, L. Laracuente, and E. V. Nordheim. Analysis of driver behavior in dilemma zones at signalized intersections. *Transportation Research Record: Journal of the Transportation Research Board*, 2030:29–39, 2007.

[9] D. Gettman and L. Head. Surrogate safety measures from traffic simulation models. Federal Highway Administation FHWA-RD-03-050, U.S. Department of Transportation, Washington D.C., 2003.

[10] P. K. Goh and Y. D. Wong. Effects of distance and speed on driver stop-versus-cross at traffic signals. *Road and Transportation Research*, 13, 2004.

[11] B.-M. Incorporated. Intersection collision avoidance study, final report. Federal Highway Administation FHWA/TX-03/4027-1, U.S. Department of Transportation, Washington D.C., 2003.

[12] S. M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. PTR Prentice Hall, New Jersey, 1993.

[13] V. L. Neale, M. A. Perez, S. E. Lee, and Z. R. Doerzaph. Investigation of driver-infrastructure and driver-vehicle interfaces for an intersection violation warning system. *Journal of Intelligent Transportation Systems*, 11(3):133–142, 2007.

[14] D. R. Ragland and A. A. Zabyshny. Intersection decision support project: Taxonomy of crossing-path crashes at intersections using GES 2000 data. No. UCB-TSC-RR-2003-08, University of California, Berkeley, 2003.

[15] H. Rakha, I. El-Shawarby, and J. Setti. Characterizing driver behavior on signalized intersection approaches at the onset of a yellow-phase trigger. *IEEE Trans. Intell. Transp. Syst.*, 8(4):630–640, Dec. 2007.

[16] D. Reid. An algorithm for tracking multiple targets. *IEEE Trans. Autom. Control*, 24(6):843–854, Dec 1979.

[17] R. A. Retting, A. F. Williams, and M. A. Greene. Red-light running and sensible countermeasures: Summary of research findings. *Transportation Research Record: : Journal of the Transportation Research Board*, (1640):23–26, 1998.

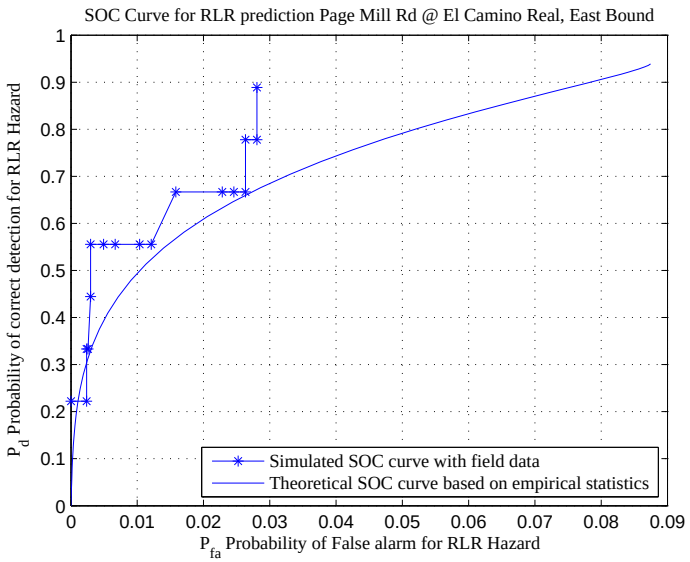
[18] Y. Sheffi and H. Mahmassani. A model of driver behavior at high speed signalized intersections. *Transportational Science*, 15:50–61, 1981.

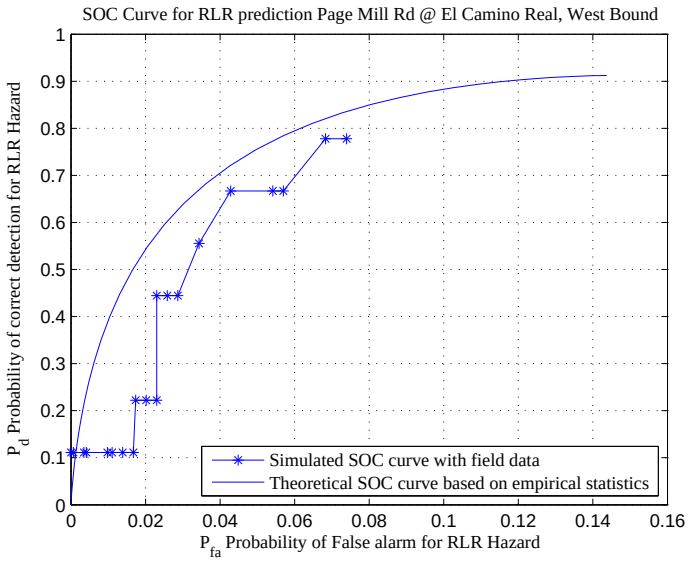
[19] R. R. Souleyrette, M. M. O'Brien, T. McDonald, H. Preston, and R. Storm. Effectiveness of all-red clearance interval on intersection crashes. Final report, Center for Transportation Research and Education,Iowa State University, Ames, Iowa, 2004.

[20] M. J. Williams. Validity of the traffic conflicts technique. *Accident Analysis and Prevention*, 13(2):133 – 145, 1981.

[21] M. Zennaro and J. A. Misener. A state map architecture for safe intelligent intersections. In *Proc. ITS America 13th Annual Meeting*, 2003.

[22] L. P. Zhang, K. Zhou, W. B. Zhang, and J. Misener. Prediction of red light running based on statistics of discrete point sensors. In *Proceedings of 88th TRB Annual Meeting*, Washington D.C., 2009.





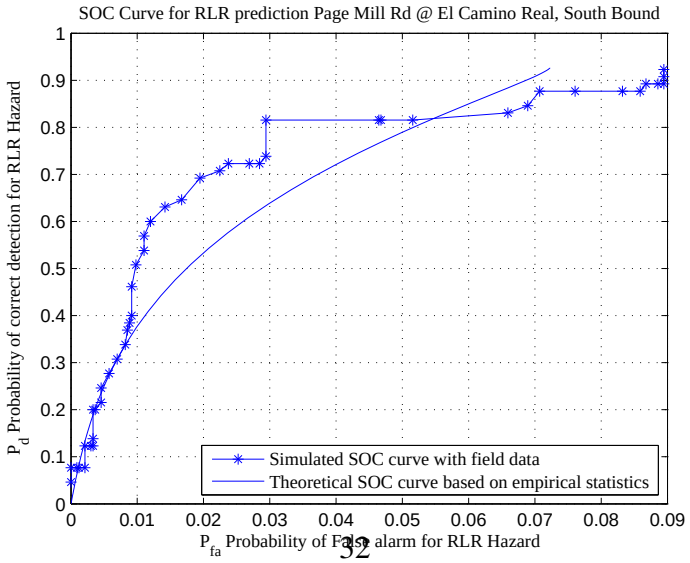


Figure 6: SOC Curves for RLR Prediction, Using Field Data

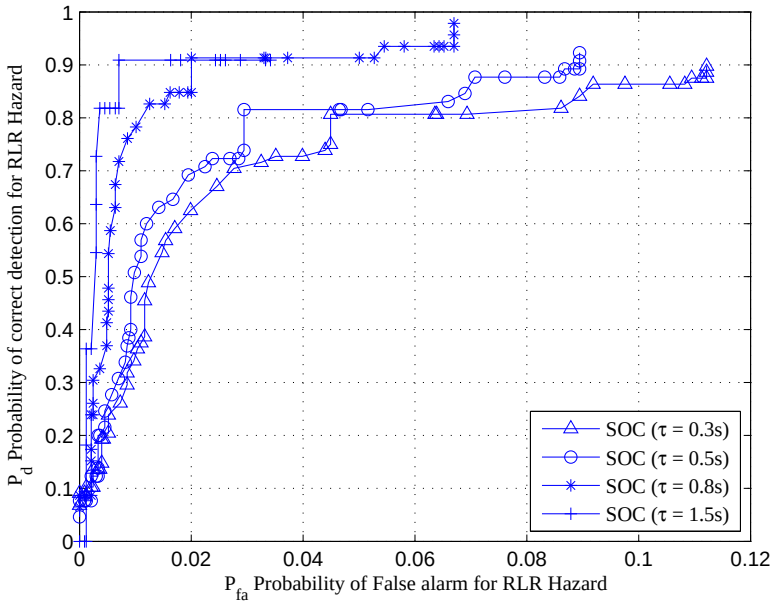


Figure 7: SOC Curves For Different Thresholds (τ)