Accident Analysis and Prevention 102 (2017) 235–247

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Estimation of red-light running frequency using high-resolution traffic and signal data

Peng Chen, Guizhen Yu*, Xinkai Wu, Yilong Ren, Yueguang Li

School of Transportation Science and Engineering, Beijing Key Laboratory for Cooperative Infrastructure System and Safety Control, Beihang University, Beijing 100191, China

ARTICLE INFO

Article history:
Received 6 December 2016
Received in revised form 17 February 2017
Accepted 11 March 2017

Keywords:
Red light running
Signalized intersection
Loop detector
High-resolution event data
Influential factors
Frequency

ABSTRACT

Red-light-running (RLR) emerges as a major cause that may lead to intersection-related crashes and endanger intersection safety. To reduce RLR violations, it's critical to identify the influential factors associated with RLR and estimate RLR frequency. Without resorting to video camera recordings, this study investigates this important issue by utilizing high-resolution traffic and signal event data collected from loop detectors at five intersections on Trunk Highway 55, Minneapolis, MN. First, a simple method is proposed to identify RLR by fully utilizing the information obtained from stop bar detectors, downstream entrance detectors and advance detectors. Using 12 months of event data, a total of 6550 RLR cases were identified. According to a definition of RLR frequency as the conditional probability of RLR on a certain traffic or signal condition (veh/1000veh), the relationships between RLR frequency and some influential factors including arriving time at advance detector, approaching speed, headway, gap to the preceding vehicle on adjacent lane, cycle length, geometric characteristics and even snowing weather were empirically investigated. Statistical analysis shows good agreement with the traffic engineering practice, e.g., RLR is most likely to occur on weekdays during peak periods under large traffic demands and longer signal cycles, and a total of 95.24% RLR events occurred within the first 1.5 s after the onset of red phase. The findings confirmed that vehicles tend to run the red light when they are close to intersection during phase transition, and the vehicles following the leading vehicle with short headways also likely run the red light. Last, a simplified nonlinear regression model is proposed to estimate RLR frequency based on the data from advance detector. The study is expected to helpbetter understand RLR occurrence and further contribute to the future improvement of intersection safety.

1. Introduction

Red-light-running (RLR)has emerged as a major cause of crashes at signalized intersections and have become a national and worldwide safety concern. According to NHTSA's Traffic Safety Facts 2008 Report (NHTSA, 2008a), there were more than 2.3 million reported intersection-related crashes, resulting in more than 7770 fatalities. Of these statistics, 762 fatalities (9.82%) and 165,000 injuries were attributable to RLR, as reported by NHTSA's Fatality Analysis Reporting System (NHTSA, 2008b). Interviews conducted with a total of 4010 drivers in the U.S. by the National Survey of Speeding and Other Unsafe Driver Actions (NHTSA, 2004) also indicated that as high as 97%of drivers feel other drivers running red lights as a major safety threat.

To reduce RLR violations and related crashes, it is critical to explore RLR exposure, identify the potential influential factors associated with RLR, estimate RLR frequency and accordingly propose engineering or enforcement countermeasures. In general, RLR is defined as a situation that approaching vehicles attempt to cross intersections during red or all-red phases. Significant research has been conducted to study RLR phenomenon (Bonneson et al., 2001). Traffic, signal timing and geometric variables have been widely identified to be associated with RLR (FHWA, 2009). The statistics of RLR frequency have also been thoroughly analyzed, mainly by accounting for the total traffic flow observed within a certain time period (Bonneson et al., 2001, 2003a). Under such definition the relationship of RLR frequency and influential factors can be statistically built and readily prepared for frequency estimation. However, the existing estimation approaches may be less accurate in view of the rare event nature of RLR and the heterogeneous characteristics

* Corresponding author.
E-mail addresses: cpeng@buaa.edu.cn (P. Chen), yugz@buaa.edu.cn (G. Yu), xinkaiwu@buaa.edu.cn (X. Wu), yilong_ren@139.com (Y. Ren), l224yg@qq.com (Y. Li).

http://dx.doi.org/10.1016/j.aap.2017.03.010
0001-4575/

of traffic flow contributing to RLR (Wang et al., 2016). For example, estimating RLR frequency with the consideration of arriving vehicles right after the onset of green phase does not make any sense.

Besides, in order to analyze RLR influential factors and frequency, it requires a large amount of data collection at intersections. Most of the existing research relies on limited off-line video detection (Gates et al., 2007; Yang and Najm, 2007), therefore fails to collect significant amount of RLR cases which need to record a large amount of video data. In contrast, loop detector data at intersections can be easily and automatically obtained. In particular, high-resolution traffic and signal event data, which enable to provide detailed vehicles' arrival and departure times against signal timing (Liu et al., 2009), can help extract sufficient RLR cases for safety assessment.

This studyaims to investigate the potential of applying event data for intersection safety research. Due to the limitation of loop detector data, the driving behavior aspects of RLR cannot be addressed. Instead, the macroscopic characteristics of RLR violations and the influential factors associated with RLR will be analyzed by leveraging the massive amount of archived high resolution event-based detector data. The research findings will be compared and verified with empirical observations on RLR in traffic engineering practice (e.g., Bonneson et al., 2001). Furthermore, to make RLR frequency estimation more objectively, RLR frequency will be estimated by redefining the base number of exposure factors such as traffic flow in appropriate analysis time intervals with RLR possibilities.

The rest of the paper is organized as follows. Section 2 presents a thorough literature review on RLR influential factors, frequency analysis and data collection methods. Section 3 offers a brief description of data preparation and then introduces an elegant method to identify RLR cases based on high-resolution event data. Data analysis results are summarized in Section 4. Influential factors of RLR are analyzed in Section 5. Section 6 proposes a simple nonlinear regression model to estimate RLR frequency. Last, we conclude this paper with some perspectives for future research.

2. Literature review

2.1. RLR influential factors

In general, exposure to RLR increases with the volume of traffic and the number of signal cycles (Baguley, 1988; Porter and England, 2000). Studies show that RLR increases with higher traffic volume, closer vehicle proximity to the intersection and higher approaching speeds (Chang et al., 1985; Mohamedshah et al., 2000; Bonneson et al., 2001). Signal timings also have significant influence on RLR. Bonneson et al. (2003a) recognized that RLR is largely affected by signal cycle length or the frequency of yellow-signal presentation. Van der Horst and Wilmink (1986) and Hagenauer et al. (1982) found that the increase of yellow and all-red intervals can help reduce RLR significantly, but Retting and Greene (1997) indicated that the increasing of all-red interval did not bring a reduction of RLR. Zegeer and Deen (1978) showed that green-extension scheme provided at intersections with actuated control would have a positive effect on reducing RLR. Furthermore, geometric variables such as approach grade (Bonneson et al., 2001), approach width, and intersection size, etc. (FHWA, 2009) may also contribute to RLR statistics.

2.2. RLR frequency analysis

The statistics of RLR frequency have been widely analyzed. In the United States, an average RLR frequency of 3 per 1000 vehicles in Arlington, Virginia (Retting et al., 1998), 36.8 per 10,000 vehicles in Fairfax, Virginia (Retting et al., 1999a), 13.2 per 10,000 vehicles in Oxnard, California (Retting et al., 1999b), 0.45-38.50 per 1000 vehicles in Iowa (Kamyab et al., 2000) and 4.1 per 1000 vehicles in Texas (Bonneson et al., 2003a) were reported. Under such a wide range of RLR frequency, one should take care when evaluating the statistics. Furthermore, RLR frequency can be estimated using identified influential factors. Bonneson and Son (2003b) developed an exponential regression model to quantify the relationship between RLR frequency and yellow interval duration, use of signal head back plates, speed, clearance path length and platoon ratio. The model was reported to be able to explain most of the systematic variation in RLR frequency in three Texas cities. Hill and Lindly (2003) developed a linear regression model to estimate RLR frequency per hour by incorporating the factors of number of lanes on the subject approach, number of lanes on the crossing approach and average daily traffic.

To the best of our knowledge, the state-of-art research has been focusing on RLR frequency analysis in terms of the unit per 1000 vehicles per hour, i.e., to estimate the frequency of RLR events by accounting for the total traffic flow observed within a certain time period. It may be less accurate in view of the rare event nature of RLR and the heterogeneous characteristics of traffic flow contributing to RLR. According to a recent study by Wang et al. (2016), most of arriving flows in the front of platoons do not have the possibility of running red lights. Thus, it does not make any sense to estimate RLR frequency with the consideration of the arriving vehicles just after the onset of green phase. On the other hand, empirical analyses indicate the arriving vehicles in a certain range around the onset of yellow are more prone to conduct RLR behavior (Bonneson and Zimmerman, 2004). In other words, drivers end up running a red light when they are trying to proceed on a yellow light in most cases, hoping they can manage to get through the intersection before the signal changes from yellow to red. Thus, it would be more reasonable to define the RLR frequency as the conditional probability of RLR on a certain traffic or signal condition, so the RLR frequency estimation will be more objective.

2.3. RLR data collection methods

To fully interpret the influential factors contributing to RLR, a large amount of RLR data plays a vital role. So far, most of research relies on data collected from video cameras. Bonneson et al. (2001) conducted a before-after study using videotape records to collect RLR data at 12 intersection approaches in three Texas cities. Gates et al. (2007) temporarily installed consumer-grade video cameras at 4 high-speed and 2 low-speed intersections in the Madison, Wisconsin to collect driver behavior data in dilemma zones. Yang and Najm (2007) examined RLR behavior using about 47,000 violation records captured by photo enforcement cameras from 11 signalized intersections in the city of Sacramento, California over a four-year period. Li and Wei (2013) modeled the dynamics of dilemma zone by collecting high-frequency vehicular trajectory data at 4 suburban high-speed signalized intersections in the state of Ohio. Al-Atawi (2014) assessed the characteristics of RLR violations in Tabuk city of the Kindom of Saudi Arabia by conducting a large-scale field survey at 38 intersections using video camera recordings. More recently, Wang et al. (2016) conducted a field investigation of RLR at 4 intersections in Shanghai, China. In these studies, video cameras were used to record the operation of vehicles entering the intersection, and then trained engineers/observers to record the information of RLR vehicles.

Video detection provides high quality data. However, this method is relatively expensive and it is difficult to collect and process video data at a large scale over a long period of time. In order to enable a view of both vehicles approaching the monitored inter-