TOWARDS A QUANTITATIVE "SAFETY" METRIC FOR AUTONOMOUS VEHICLES

Arturo Tejada, Michel J. E. Legius
TNO, Integrated Vehicle Safety Department
5700 AT Helmond The Netherlands

Paper Number 19-0012

ABSTRACT

Future mobility systems are expected to incorporate a broad range of transport modalities (passenger cars, truck platoons, etc.) at different automation levels (SAE Levels 3/4/5). During operation, automated vehicles will have to independently take safety-critical decisions (e.g., when to brake or change lanes) and estimate the impact of their behavior on the surrounding traffic, thus balancing individual and group safety. To achieve this, automated vehicles will require a quantitative metric of safety to guide their actions.

This article proposes one such metric, suitable for decision-making and autonomous navigation. The metric is meant to provide a quantification of the risk a vehicle incurs during operation by taking into account three main aspects of its operation: the probability of a hazard occurring (e.g., a rear-end collision), the potential impact of the driving conditions on the health of the vehicle's passengers were the hazard to occur, and the capability of the vehicle to avoid the hazard. The article focuses on introducing the conceptual aspects of the metric first and then presents the initial results on estimating and collision probabilities. The other two aspects will be addressed elsewhere.

INTRODUCTION

Automated vehicles (car, buses, trucks) are widely considered to be a promising solution to the road safety and congestion problems. On the one hand, such vehicles would not be subject to distractions or lapses of judgement that currently lead to the majority of road accident involving human drivers [1], thus improving road safety. On the other hand, their shorter reaction times and ability to communicate with road infrastructure would allow them to drive more efficiently on the roads, increasing road utilization and capacity, while improving traffic flow.

To realize the aforementioned benefits, automated vehicles would have to drive according to a notion of *correct behavior*[1], which presumably would maximize (or at least maintain) the safety level of all traffic participants. Trained (and experienced) human drivers are able to judge the safety of their situations and act accordingly (most of the time). This is generally done unconsciously based on implicitly learned behaviors and models of the world. Automated vehicles, on the other hand, must make explicit judgements about their safety in order to take decisions regarding their behaviors[2]. Such decisions may include when to activate automated emergency braking systems, which trajectory to follow while driving, which route to use to reach a particular location, when to allow a human driver to regain vehicle control, etc. Thus, reasoning about safety would be needed at several layers of an automated vehicle's architecture [2, 3].

Much as the notion of correct behavior (see Footnote 1), safety is also a difficult concept to define precisely, and it varies with the stake holder considering it. As summarized in the next section, safety has been addressed from at least three perspectives: traffic system safety, vehicle hazard (collision) avoidance, and vehicle certification (functional safety). Briefly, from the traffic system perspective, safety is related to understanding (and preventing) the factors that contribute to traffic crashes and injuries (e.g., vehicle technologies, infrastructure design, etc. [4,5]); safety from the hazard avoidance perspective is related to determining thresholds on proximity metrics to trigger warning and/or collision avoidance systems (see, among

[1] *Correct behavior* is a societally-agreed concept that encompass not only normative or engineering-like goals (e.g., "follow the traffic rules") but also elements such as "use an acceptable driving style", which vary widely by country, age group, etc.

[2] Arguably, an automated vehicle could be completely controlled by machine-learning-based algorithms trained to mimic human driving behavior without explicitly reasoning about safety. Not all automated vehicles, however, would be so controlled.

many examples, [6, 7]; while safety from the certification perspective is related to developing the vehicular electronics and software at the necessary automotive safety integrity levels (ASILs) [8]. Unfortunately, none of these approaches directly provide a definition of safety that could be operationalized to reason safety as needed for autonomous vehicle operation.

This article, introduces a framework to compute one such metric. It can be thought of as an extension of existing techniques on estimation of probability of collision [9, 10] to incorporate measures of both the consequences of a potential collision and the ability of the vehicle to avoid such collision. The framework is modelled after the concepts used in the ISO 262622 standard to assigned ASIL levels, and the resulting metric can be considered as a measure of interaction severity, when applied to interactions of two vehicles.

The rest of this article is organized as follows: The framework of estimating safety is introduced next, after a brief summary of available literature on safety definitions. This is followed by a description of our initial work on estimating the probability of collision in two dimensions, one of the main three elements on our framework (the other two elements are outlined but described in detail elsewhere). Finally, the paper ends with out conclusions.

FRAMEWORK FOR SAFETY ESTIMATION

As mentioned in the introduction, the framework for safety estimation proposed here draws from concepts of safety defined by other researchers and stakeholders. To avoid confusion between the use of "safety" in the vernacular, and "safety" as a subject of study, the latter will be italicized in the sequel. Several concepts of *safety* are summarized next.

A Brief Overview of Safety in the Literature

Perhaps the most frequently recognized concept of *safety* comes from the field of traffic safety: it is the absence of road accidents that may lead to severe injury, deaths, and or property damages (see, e.g., [11]). From this perspective, *safety* cannot be directly measured[3]. The *lack of safety*, however, can be described via accident statistics, and it is know to have a large impact in terms of loss of life, livelihood, and economic output [13]. This is why its minimization is of great interest for governments and regulators. Minimizing (or, hopefully, eliminating) the *lack of safety*, however, depends not only on the actions of individual vehicles in traffic, but also on the right combination of vehicle technologies, infrastructure design, and traffic policies. Thus, the aim of traffic safety research is not to measure *safety* in order to allow vehicle automation, but rather to show whether or not the introduction of new technologies or traffic policies lead to demonstrable reductions of severe road accident statistics (see [14, 15] and the references therein).

The aim of minimizing the *lack of safety* as defined above, gave raise to a number of technologies that help vehicles avoid or minimize the effect of severe traffic conflicts, which in turn can lead to severe collisions [14]. Severe traffic conflicts can be defined as traffic interactions where two or more participants are in collision course and "too close" either in space or time [16]. Collision courses are determined using simplifying assumptions on the vehicles' behavior (e.g., constant acceleration), while proximity is measured using any number of indicators, most commonly, time to collision. Appropriate thresholds over these indicators allow for the automatic activation of a vehicle's warning and/or collision avoidance systems [6, 7]. Vehicles equipped with such technologies are assumed to be safer because they minimize the chance of occurrence and/of the effects of accidents, which is usually confirmed via testing (see, e.g., [17]). Note, however, that *safety* does not need to be (and it is not) directly defined/measured in these approaches.

A different perspective altogether on automotive safety is that of *functional safety*. This concepts, presented in detail in the ISO26262 standard [8] is related to minimizing the risk of occurrence of particular hazards due to technological failures (e.g., rear-end collision due to braking failure). To accomplish this, a vehicle's hardware (i.e., electronics) and software are developed in a way that they attain specific automotive safety integrity levels (ASILs). These, in turn, are derived by analyzing the time and possibility of the vehicle

[3]From this perspective, *safety* (i.e., the absence of loss of life under all traffic circumstances) cannot be formally proven. This is, however, a very compelling and important goal to pursue, as stated by the Vision Zero philosophy [12]

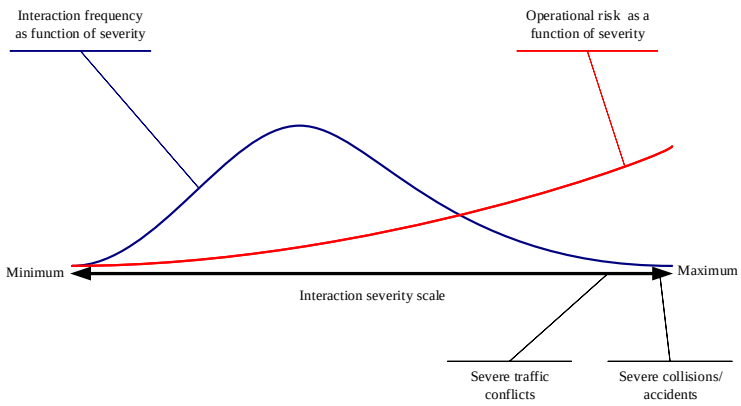

Figure 1: Severity scale for traffic interactions and associated risk (adapted from [4])

being exposed to the said hazards, the gravity of injuries that might be caused were the hazards to occur, and the likelihood that the injuries could be prevented by the actions of a typical driver. A vehicular system so designed is said to be functionally safe. Though functional safety is precisely defined in the ISO26262 standard, it cannot be directly quantified.

Finally, it is important to point out that there are several metrics in the literature that are used to allow vehicle automation [15, 18, 19]. Such metrics have been used for, among other applications, collision avoidance and path planning [6, 7, 20]). To the best of our knowledge, such metrics are application specific and do not always take into account the consequence of potential collisions on the drivers involved. The framework proposed next aims to address both these issues.

On Safety Estimation

A well-known concept from traffic safety literature is that all traffic interactions can be placed along a continuous "severity" scale [4, 14] (see Figure 1). On one end of the scale lie the least severe interactions, which lead to accidents with very low probability (e.g., a vehicle driving alone in a road). Towards the middle of the scale lie the mid-severe interactions. They are the most frequent interactions and, generally speaking, lead to accidents with low probability. On the other end of the scale one finds the most severe and rare interactions (i.e., the severe traffic conflicts) that lead to severe accidents with (presumably) high probability. This suggests that one could reason about the degree of *safety* of a given vehicle in traffic, by quantifying the severity of the traffic interactions affecting it at a given time.

Unfortunately, there seems not to be yet a consensus in the literature on how to quantify interaction severity, although it is expected that multiple indicators should be combined to produce such quantification (see, e.g., [14, 19, 21]). Here, we propose an approach motivated by the method ISO 26262 standard to assign ASILs. That is, each traffic interaction involving a vehicle of interest (called *host*) is treated as a potential hazard and assigned a severity value equal to the operational risk it imposes on the host. Clearly, the higher the severity of the interaction, the higher the risk for the host (see Figure 1).

To assign a risk value, we consider the potential hazard's likelihood, hazardousness, and avoidability[4]. The relationship among these factors are illustrated in Fogire 2: the (blue) host vehicle keeps a constant distance with the (yellow) vehicle in front of it. During this interaction, another (green) vehicle begins a cut-in maneuver between them. This maneuver creates a potential collision hazard. The risk induced by this hazard depends on the likelihood of the collision (usually estimated from prediction models). Intuitively, the higher the collision likelihood, the higher the host's risk. Further, a mild collision presents a lower risk to the passengers involved than a severe collision , as the latter has a higher chance to produce severe injuries. Thus,

[4]These concepts mirror exposure, severity and controllability in ASIL assignment [8].

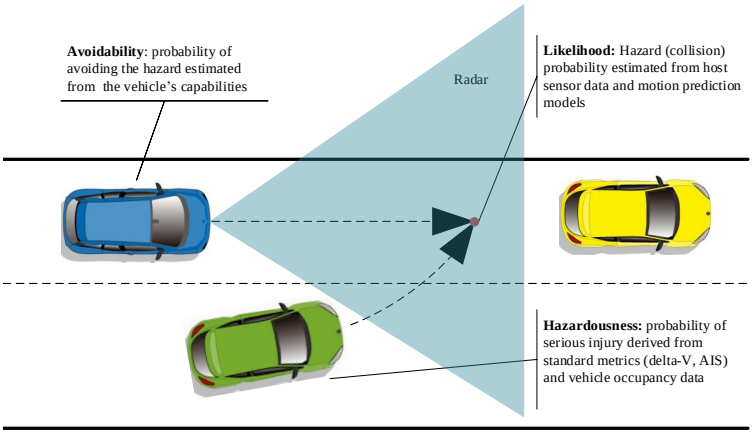

Figure 2: Example scenario showing the factors contributing to operational risk.

the hazardousness of the potential collision has also a strong bearing on the host's risk. Finally, depending on its capabilities, the host may be able to ameliorate the hazard's consequences or to avoid it all together. Thus, the hazard's avoidability also plays an important role on understanding the host's operational risk.

Formally, let $S : \mathbb{R}^+ \to [0, 1]$ and $R : \mathbb{R}^+ \to [0, 1]$ denote the host's (degree of) safety and operational risk functions respectively ($\mathbb{R}^+$ denotes the non-negative real numbers). Then, the host's (degree of) safety in the next $T \in \mathbb{R}^+$ seconds is given by

$$S(T) = 1 - R(T) \triangleq \mathrm{P}\{\mathcal{H}\}(1 - \mathrm{P}\{\mathcal{A}\})\mathrm{P}\{\mathcal{I}|\mathcal{H}\}, \tag{1}$$

where $\mathrm{P}\{\mathcal{H}\}$ denotes the probability that "a hazard occurs in the next T seconds" and $\mathrm{P}\{\mathcal{I}|\mathcal{H}\}$ denotes the probability that "severe injury or death occur if the hazard occurs in the next T seconds" and $\mathrm{P}\{\mathcal{A}\}$ denotes the probability that the vehicle "can perform an action to avoid the hazard in the next T seconds". These three probabilities denote, respectively, the likelihood, hazardousness and avoidability of hazard $\mathcal{H}$.

Remark 1. That a hazard occurs in a given period depends, among other factors, on the actions of the host and other traffic participants during that period. An automated vehicle can only *estimate* these future actions based on past and present sensor data and behavior prediction models. The greater the period T, the more uncertain the estimations become. This uncertainty is captured by $\mathrm{P}\{\mathcal{H}\}$. ♦

Remark 2. $\mathrm{P}\{\mathcal{A}\}$ estimates the capability of the host to perform a hazard avoidance action on time to prevent the hazard from occurring. In (1), as a first approximation, it is assumed that $\{\mathcal{H}\}$ and $\{\mathcal{A}\}$ are independent events. However, it is clear that the more capable the host is of avoiding a hazard, the lower the latter's likelihood. Nevertheless, note that even when $\mathrm{P}\{\mathcal{A}\} = 1$, $\mathrm{P}\{\mathcal{H}\} > 0$ since the hazard likelihood does not depend solely on the hosts actions. ♦

Finally, if the host faces more than one hazard simultaneously (see Figure 2), (1) can be extended as follows

$$S(T) = 1 - R(T) \triangleq \sum_i \mathrm{P}\{\mathcal{H}_i\}(1 - \mathrm{P}\{\mathcal{A}_i\})\mathrm{P}\{\mathcal{I}_i|\mathcal{H}_i\}, \tag{2}$$

where each $\mathcal{H}_\rangle$, $i = 1, 2, \cdots$, denotes a separate independent hazard (under the assumption that hazards can be treated independently).

The rest of the document will focus only on collision hazards and will present a method to estimate $P\{\mathcal{H}\}$. Details on how to estimate the probability of injury given specific collision conditions can be found in [22]. Methods to estimate collision avoidance capabilities are currently under investigation and will be presented elsewhere.

ON PROBABILITY OF COLLISION

As mentioned in the previous section, $\mathrm{P}\{\mathcal{H}\}$ represents the host's uncertainty in determining whether a collision will occur in a give timeframe. It arises due to: 1) imprecisions in the host's sensor measurements and 2) the uncertainty associated with predicting the *future* behavior of other traffic participants (called *targets*) relative to the host.

The general approach used to estimate $\mathrm{P}\{\mathcal{H}\}$ is as follows:

- The host's perception system measure or derives kinematic quantities like host/target positions, headings, lengths and widths, etc.
- These quantities, together with behavioral assumptions on both the host and target, are used to infer the relative positions of the host and target in the future T seconds[5].
- Determine if the target and host overlap.

Typical behavioral assumptions used to estimate future host/target behavior are that they either move with constant speed or with constant acceleration (see [9]). This allows one estimate both their future relative positions and their associated probability density functions. From this information one can infer the probability density function of the host/target overlap. Determining whether two vehicles overlap is akin to determining whether two convex polygons intersect. This is discussed next.

On Intersections of Convex Polygons

The main tool used to analyze whether two convex sets intersect is called the Separating Hyperplane Theorem (SHT, see [23]).

Theorem 1 (SHT)**.** *Suppose C and D are nonempty disjoint convex sets, i.e., $C \cap D = \phi$. Then, there exists $a \neq 0$ and b such that $a^T x \leq b$ for all $x \in C$ and $a^T x \geq b$ for all $x \in D$.*

The proof of this theorem is constructive (see [23, p. 46]) and can be used to extend the theorem in several ways by adding additional conditions to the sets C and D as shown next (see [24, 25]):

Lemma 2. *Suppose $C, D \in \mathbb{R}^n$ are nonempty, closed, convex sets, at least one of which is bounded, and are such that $C \cap D = \phi$. Then, there exists $a \neq 0$ and b such that $a^T x < b$ for all $x \in C$ and $a^T x > b$ for all $x \in D$.*

The converse of these theorems seems to be true for finite dimensional spaces (like $\mathbb{R}^n$), though no formal proof has been found. In the finite dimensional case, using the additional concept of "separating axis" (i.e., a line perpendicular to a separating hyperplane) the following corollary of the Separating Hyperplane Theorem can be stated:

Lemma 3 (Separating Axis Theorem (SAT))**.** *Let $C, D \in \mathbb{R}^n$ be nonempty, closed, convex sets. If there exists a line L for which the projections of C and D, respectively $P_L(C)$ and $P_L(D)$, onto it do not intersect, then the $C \cap D = \phi$.*

SAT is stated without proof by most authors (although a proof is reportedly available in [26]). This corollary does not show how to find L so one presumably would have to identify it by inspection or by trial and error. Fortunately, in the case of convex polygons, the search space for L can be narrowed significantly. To do this, let L be defined as follows:

$$L \triangleq \{x \in \mathbb{R}^2 | x = a + t\hat{v}, \quad t \in (-\infty, \infty)\},$$

[5] More precisely, one should estimate the positions of the host and the target at time $\min\{TTC, T\}$, where TTC denotes the time to collision [15]

where a is a point in $\mathbb{R}^2$ and $\hat{v}$ is a unitary vector in the desired direction of L. Further, let the operator P_L be defined as follows:

$$\begin{aligned} P_L : \mathbb{R}^2 &\to \mathbb{R}^2 \\ x &\mapsto a + \langle x - a, \hat{v} \rangle \hat{v}, \end{aligned}$$

where $\langle ., . \rangle$ denotes the standard inner product. $P_L(x)$ returns the location of the orthogonal projection of x on L. This concept can be extended to a set $C \subset \mathbb{R}^2$, with a slight abuse of notation, as follows:

$$P_L(C) \triangleq \{y \in R^2 | y = P_L(x), x \in C\}$$

The next result then follows (see [27, sec. 7.7.2]).

Corollary 4 (SAT for Convex Polygons)**.** *Let $C, D \in \mathbb{R}^n$ be $n-$ and $m-$sided convex polygons, respectively, and let S be a set of lines, each normal to a different edge of C and D. If there exist $L \in S$ such that $P_L(C) \cap P_L(D) = \phi$ then $C \cap D = \phi$.*

Note that S is not unique, since the translation of a separating axis is also a separating axis. Hence, it is sufficient for S to contains lines that cross the origin (i.e., for which $a = 0$). Also note that $P_L(C)$ and $P_L(D)$ are, by construction, line segments.

Detection of Vehicle Collision

To apply the results of the previous subsection to determining whether two vehicle overlap (i.e., have collided), consider the setup in Figure 3, which shows a host and a target. All coordinates in this figure are measured with respect to an arbitrary, fixed, ground coordinate system $\{G\}$. The figure's nomenclature is given in Table 1.

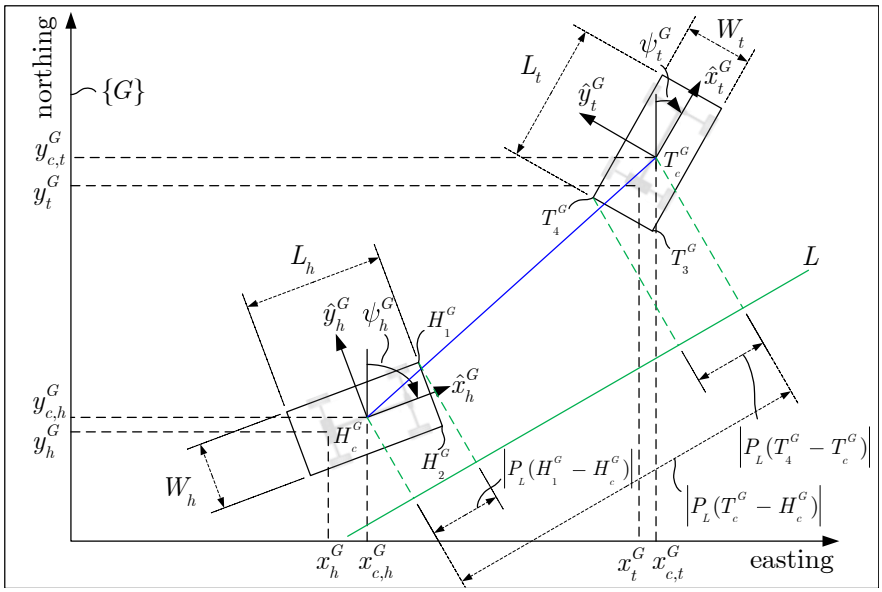

Figure 3: Coordinate framework for collision detection showing a host and a target vehicle.

To determine whether the host and target overlap, one can apply the SAT for convex polygons. Thus, assume that in Figure 3 the line L is a separating axis. Theorem 1 and Lemma 3 imply that

$$P_L(\text{Host}) \cap P_L(\text{Target}) = \phi \iff |P_L(T^G_c - H^G_c)| > 1/2|P_L(\text{Host})| + 1/2|P_L(\text{Target})|. \tag{3}$$

That is, the host and the target *do not collide* if the right hand side on the expression above holds[6]. The latter can be expanded as follows:

$$|P_L(T_c^G - H_c^G)| > \max\{|P_L(H_1^G - H_c^G)|, |P_L(H_2^G - H_c^G)|\} + \max\{|P_L(T_4^G - T_c^G)|, |P_L(T_3^G - T_c^G)|\}. \quad (4)$$

This inequality can be expressed in term of the variables shown in Figure 3 by noticing that

$$|P_L(H_1^G - H_c^G)| = |\langle H_1^G - H_c^G, \hat{v}\rangle \hat{v}| = |\langle L_h/2\hat{x}^H + W_h/2\hat{y}^H, \hat{v}\rangle|,$$

and that

$$\begin{aligned} |P_L(H_2^G - H_c^G)| &= |\langle L_h/2\hat{x}^H - W_h/2\hat{y}^H, \hat{v}\rangle|, \\ |P_L(T_4^G - T_c^G)| &= |\langle -L_t/2\hat{x}^T + W_t/2\hat{y}^T, \hat{v}\rangle|, \\ |P_L(T_3^G - T_c^G)| &= |\langle -L_t/2\hat{x}^T - W_t/2\hat{y}^T, \hat{v}\rangle|. \end{aligned}$$

Thus, condition (4) is equivalent to

$$\begin{aligned} |P_L(T_c^G - H_c^G)| > \max\{|\langle L_h/2\hat{x}^H + W_h/2\hat{y}^H, \hat{v}\rangle|, |\langle L_h/2\hat{x}^H - W_h/2\hat{y}^H, \hat{v}\rangle|\} \\ + \max\{|\langle L_t/2\hat{x}^T + W_t/2\hat{y}^T, \hat{v}\rangle|, |\langle L_t/2\hat{x}^T - W_t/2\hat{y}^T, \hat{v}\rangle|\}. \quad (5) \end{aligned}$$

According to Corollary 4, it is sufficient to limit the search for a line L that would fulfill (5) to those parallel to the sides of the host and target vehicles. If no one if these lines satisfies (5), then the host and target vehicles overlap. This leads to the following result.

Corollary 5. *The host and target vehicles shown in Figure 3* do not *collide if and only if* any *the following four conditions holds:*

$$\begin{aligned} |\langle T_c^G - H_c^G, \hat{x}_h^G\rangle| &> L_h/2 + \max\{|\langle L_t/2\hat{x}_t^G + W_t/2\hat{y}_t^G, \hat{x}_h^G\rangle|, |\langle L_t/2\hat{x}_t^G - W_t/2\hat{y}_t^G, \hat{x}_h^G\rangle|\}, \\ |\langle T_c^G - H_c^G, \hat{y}_h^G\rangle| &> W_h/2 + \max\{|\langle L_t/2\hat{x}_t^G + W_t/2\hat{y}_t^G, \hat{y}_h^G\rangle|, |\langle L_t/2\hat{x}_t^G - W_t/2\hat{y}_t^G, \hat{y}_h^G\rangle|\}, \\ |\langle T_c^G - H_c^G, \hat{x}_t^G\rangle| &> L_t/2 + \max\{|\langle L_h/2\hat{x}_h^G + W_h/2\hat{y}_h^G, \hat{x}_t^G\rangle|, |\langle L_h/2\hat{x}_h^G - W_h/2\hat{y}_h^G, \hat{x}_t^G\rangle|\}, \\ |\langle T_c^G - H_c^G, \hat{y}_t^G\rangle| &> W_t/2 + \max\{|\langle L_h/2\hat{x}_h^G + W_h/2\hat{y}_h^G, \hat{y}_t^G\rangle|, |\langle L_h/2\hat{x}_h^G - W_h/2\hat{y}_h^G, \hat{y}_t^G\rangle|\}. \end{aligned}$$

Proof : This result follows directly from Theorem 1 and Corollary 4. ∎

The inequalities in Corollary 5 can be simplified by using a more conservative definition of "non-collision" conditions as follows.

Corollary 6. *The host and target vehicles shown in Figure 3* do not *collide if* any *the following four conditions holds:*

$$|\langle T_c^G - H_c^G, \hat{x}_h^G\rangle| > \left(L_h + \sqrt{(L_t)^2 + (W_t)^2}\right)/2, \quad (6)$$

$$|\langle T_c^G - H_c^G, \hat{y}_h^G\rangle| > \left(W_h + \sqrt{(L_t)^2 + (W_t)^2}\right)/2, \quad (7)$$

$$|\langle T_c^G - H_c^G, \hat{x}_t^G\rangle| > \left(L_t + \sqrt{(L_h)^2 + (W_h)^2}\right)/2, \quad (8)$$

$$|\langle T_c^G - H_c^G, \hat{y}_t^G\rangle| > \left(W_t + \sqrt{(L_h)^2 + (W_h)^2}\right)/2. \quad (9)$$

Although simpler, conditions (6)-(9) are more conservative. That is, a host and a target may not satisfy these condition and still not be in collision. As it will be seen next, these leads to an over estimation of the probability of collision.

[6]The inequality in (3) is often stated as a "greater or equal" inequality, to allow for the fact that the vehicles may share common edge (i.e., just touch). Here, a "more than" inequality is used to guarantee full separation between vehicles.

[1]The world currently outputs the position of the center of a vehicle's back axel (host or target). Here it is assumed that, in the future, the world model will also output the length, L, and width, W, every target vehicle.

Table 1: Figure 3 nomenclature.

	Signal	Explanation	Unit
Target	$T_c^G = (x_{c,t}^G, y_{c,t}^G)$	Target geometric center position w.r.t. $\{G\}$	m,m
	$x_{c,t}^G$	Target geometrical center[1] easting position	m
	$y_{c,t}^G$	Target geometrical center northing position	m
	x_t^G	Target rear axle center easting position	m
	y_t^G	Target rear axle center northing position	m
	ψ_t^G	Target rear axle heading (clockwise positive)	rad
	$\hat{x}_t^G$	Unitary vector parallel to the Target's heading in $\{G\}$ coordinates	
	$\hat{y}_t^G$	Unitary vector perpendicular to the Target's heading in $\{G\}$ coordinates	
	L_t	Target vehicle's length	m
	W_t	Target vehicle's width	m
	T_3^G (T_4^G)	Target vehicle's back right (left) corner position w.r.t. $\{G\}$	m,m
Host	$H_c^G = (x_{c,h}^G, y_{c,h}^G)$	Host geometric center position w.r.t. $\{G\}$	m,m
	$x_{c,h}^G$	Host geometrical center[1] easting position	m
	$y_{c,h}^G$	Host geometrical center northing position	m
	x_h^G	Host rear axle center easting position	m
	y_h^G	Host rear axle center northing position	m
	ψ_h^G	Host rear axle heading (clockwise positive)	rad
	$\hat{x}_h^G$	Unitary vector parallel to the Host's heading in $\{G\}$ coordinates	
	$\hat{y}_h^G$	Unitary vector perpendicular to the Host's heading in $\{G\}$ coordinates	
	L_h	Host vehicle's length	m
	W_h	Host vehicle's width	m
	H_1^G (H_2^G)	Host vehicle's front right (left) corner position w.r.t. $\{G\}$	m,m
Projections	L	Separating axis candidate	
	$\lvert P_L(T_c^G - H_c^G)\rvert$	Projection magnitude of vector $T_c^G - H_c^G$ over L	m
	$\lvert P_L(H_1^G - H_c^G)\rvert$	Projection magnitude of vector $H_1^G - H_c^G$ over L	m
	$\lvert P_L(T_4^G - T_c^G)\rvert$	Projection magnitude of vector $T_4^G - T_c^G$ over L	m

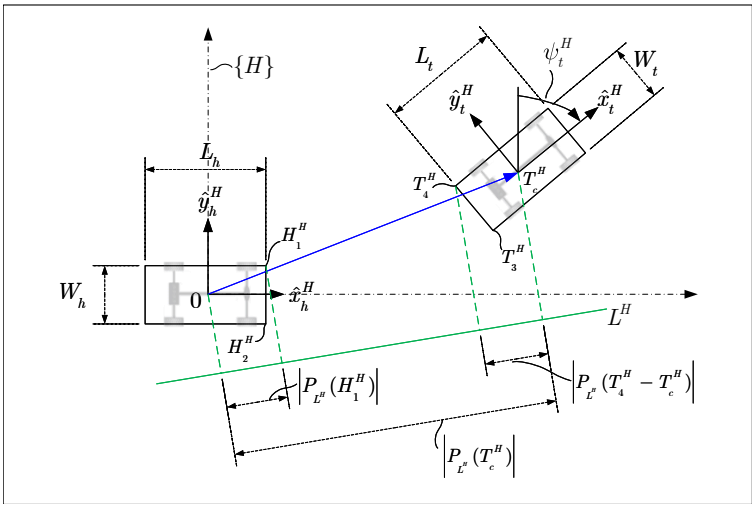

Figure 4: Host coordinate framework for collision detection showing a host and a target vehicle. The variables are defined as in Table 1, replacing the superscript ^{G} with ^{H}.

Estimation of Probability of Collision

The estimation of probability of collision is performed by the host vehicle based on measurements provided by its perception system. These measurements include the target position w.r.t. the host, T_c^H, and the target's heading w.r.t., ψ_t^H, and assume the frame of reference shown in Figure 4 (a rotated version of Figure 3). It follows from this figure that:

$$H_c^H = 0, \quad \hat{x}_h^H = (1,0), \quad \hat{y}_h^H = (0,1), \quad \hat{x}_t^H = (\sin(\phi_t^H), \cos(\phi_t^H)), \quad \hat{y}_t^H = (-\cos(\phi_t^H), \sin(\phi_t^H)).$$

As mentioned before, the uncertainty associated with the host vehicle sensor measurements and the predictions of future host/target behavior is what makes checking (6)-(9) non-deterministic. The uncertain sensor measurements can be treated as random variables with associated Probability Density Functions (PDFs). These variables will be denoted in the sequel in boldface fonts. The primary sensor measures, upon which any other quantities are derived, are listed next.

Assumption 1. *The following (primary) quantities are assumed to be independent random variables defined over the probability space* $(\Omega, \mathscr{F}, \mathrm{P})$. *They are shown with their associated PDFs.*

$$\boldsymbol{T}_c^H = (\boldsymbol{x}_{c,t}^H, \boldsymbol{y}_{c,t}^H) \sim f_{T_c} = f_{x_{c,t}} f_{y_{c,t}}, \quad \text{where} \quad \boldsymbol{x}_{c,t}^H \sim f_{x_{c,t}} \quad \text{and} \quad \boldsymbol{y}_{c,t}^H \sim f_{y_{c,t}}$$
$$\boldsymbol{\psi}_t^H \sim f_{\psi_t}, \quad \boldsymbol{L}_t \sim f_{L_t}, \quad \boldsymbol{W}_t \sim f_{W_t}.$$

Since current automated vehicle not in general have the ability of directly estimating W_t or L_t from sensor measurements, the following assumption will be accepted as true in the sequel.

Assumption 2. *The target's size is communicated via vehicle-to-vehicle communication to the host.*

These assumptions and conditions (6)-(9) can be used to derive an over-estimate of the probability of collision as follows: Let $C_1, C_2, C_3, C_4 \subset \Omega$ be the following events

$$C_1 \triangleq \{\omega \in \Omega \mid |\boldsymbol{x}_{c,t}^H| > \left(L_h + \sqrt{(L_t)^2 + (W_t)^2}\right)/2\},$$
$$C_2 \triangleq \{\omega \in \Omega \mid |\boldsymbol{y}_{c,t}^H| > \left(W_h + \sqrt{(L_t)^2 + (W_t)^2}\right)/2\},$$
$$C_3 \triangleq \{\omega \in \Omega \mid |\boldsymbol{x}_{c,t}^H \sin(\boldsymbol{\psi}_t^h) + \boldsymbol{y}_{c,t}^H \cos(\boldsymbol{\psi}_t^h)| > \left(L_t + \sqrt{(L_h)^2 + (W_h)^2}\right)/2\},$$
$$C_4 \triangleq \{\omega \in \Omega \mid |\boldsymbol{x}_{c,t}^H \cos(\boldsymbol{\psi}_t^h) - \boldsymbol{y}_{c,t}^H \sin(\boldsymbol{\psi}_t^h)| > \left(W_t + \sqrt{(L_h)^2 + (W_h)^2}\right)/2\}.$$

Further, let $A \subset \Omega$ be the event "host and target do not collide" defined according to Corollary 6. That is, $A \triangleq \left\{\omega \in \Omega \,|\, \omega \in \bigcup_{i=1}^{4} C_i\right\}$. It follows from this definition and from De Morgan's laws that $\bar{A} \triangleq \Omega - A$, the "host and target collide" event, is given by:$\bar{A} \triangleq \left\{\omega \in \Omega | \omega \in \bigcap_{i=1}^{4} \bar{C}_i\right\}$, where $\bar{C}_i = \Omega - C_i$, $i = 1, \cdots, 4$. This lead to the following result

Theorem 7. *Consider the host and target vehicle in Figure 4. Under Assumptions 1 and 2, the probability of collision between the host and target,* $\mathrm{P}\{\mathcal{H}\}$ *can be* overestimated *as follows:*

$$\mathrm{P}\{\mathcal{H}\} = \mathrm{P}\{\text{"Host-Target Collide"}\} \leq \mathrm{P}\{\omega \in \Omega \,|\, \omega \in \bar{A}\} = \mathrm{P}\{(\boldsymbol{x}_{c,t}^H, \boldsymbol{y}_{c,t}^H, \boldsymbol{\psi}_t^H) \in R_1 \cap R_2(\boldsymbol{\psi}_t^H) \times [0, 2\pi)\},$$

where $R_1, R_2(\boldsymbol{\psi}_t^H) \subset \mathbb{R}^2$ *are rectangular regions in* $\mathbb{R}^2$ *given, respectively, by:*

$$R_1 \triangleq \{(x,y) \in \mathbb{R}^2 \,|\, |x| \leq 0.5(L_h + \sqrt{(L_t)^2 + (W_t)^2}), |y| \leq 0.5(W_h + \sqrt{(L_t)^2 + (W_t)^2})\},$$

and

$$R_2(\boldsymbol{\psi}_t^H) \triangleq \left\{(x,y) \in \mathbb{R}^2 \,|\, (x,y) = r \begin{bmatrix} \sin(\boldsymbol{\psi}_t^H) \\ \cos(\boldsymbol{\psi}_t^H) \end{bmatrix} + s \begin{bmatrix} \cos(\boldsymbol{\psi}_t^H) \\ -\sin(\boldsymbol{\psi}_t^H) \end{bmatrix}; \right.$$
$$\left. |r| \leq 0.5(L_t + \sqrt{(L_h)^2 + (W_h)^2}), |s| \leq 0.5(W_t + \sqrt{(L_h)^2 + (W_h)^2}) \right\}.$$

Proof: Recall that A is defined based on conditions (6)-(9). Since these are conservative, it follows that A is a subset of the event "Host and Target do not Collide". This in turn implies that $\mathrm{P}\{\text{"Host-Target Collide"}\} \leq \mathrm{P}\{\omega \in \Omega \,|\, \omega \in \bar{A}\}$. Next, observe from the definition of A that

$$\bar{A} = \left\{\omega \in \Omega | \omega \in (\bar{C}_1 \cap \bar{C}_2) \cap (\bar{C}_3 \cap \bar{C}_4) \cap \Omega\right\}. \tag{10}$$

The three terms in the RHS of the expression above can be further developed.

$$\begin{aligned} \bar{C}_1 \cap \bar{C}_2 &= \left\{\omega \in \Omega \,|\, |\boldsymbol{x}_{c,t}^H| \leq \left(L_h + \sqrt{(L_t)^2 + (W_t)^2}\right)/2\right\} \\ &\qquad \bigcap \left\{\omega \in \Omega \,|\, |\boldsymbol{y}_{c,t}^H| \leq \left(W_h + \sqrt{(L_t)^2 + (W_t)^2}\right)/2\right\} \\ &= \left\{\omega \in \Omega \,|\, |\boldsymbol{x}_{c,t}^H| \leq \left(L_h + \sqrt{(L_t)^2 + (W_t)^2}\right)/2, |\boldsymbol{y}_{c,t}^H| \leq \left(W_h + \sqrt{(L_t)^2 + (W_t)^2}\right)/2\right\} \\ &= \left\{\omega \in \Omega \,|\, (\boldsymbol{x}_{c,t}^H, \boldsymbol{y}_{c,t}^H) \in R_1\right\}. \end{aligned} \tag{11}$$

Similarly:

$$(\bar{C}_3 \cap \bar{C}_4) \cap \Omega = (\bar{C}_3 \cap \bar{C}_4) \cap \{\omega \in \Omega \,|\, \boldsymbol{\psi}_t^H \in [0, 2\pi)\}$$

and

$$(\bar{C}_3 \cap \bar{C}_4) = \{\omega \in \Omega \,|\, |\boldsymbol{x}_{c,t}^H \sin(\boldsymbol{\psi}_t^h) + \boldsymbol{y}_{c,t}^H \cos(\boldsymbol{\psi}_t^h)| \leq \left(L_t + \sqrt{(L_h)^2 + (W_h)^2}\right)/2,$$
$$|\boldsymbol{x}_{c,t}^H \cos(\boldsymbol{\psi}_t^h) - \boldsymbol{y}_{c,t}^H \sin(\boldsymbol{\psi}_t^h)| \leq \left(W_t + \sqrt{(L_h)^2 + (W_h)^2}\right)/2\}.$$

The above expression can be further developed by letting $\boldsymbol{r} = \boldsymbol{x}_{c,t}^H \sin(\boldsymbol{\psi}_t^h) + \boldsymbol{y}_{c,t}^H \cos(\boldsymbol{\psi}_t^h)$ and $\boldsymbol{s} = \boldsymbol{x}_{c,t}^H \cos(\boldsymbol{\psi}_t^h) - \boldsymbol{y}_{c,t}^H \sin(\boldsymbol{\psi}_t^h)$, and observing that

$$\begin{bmatrix} \boldsymbol{r} \\ \boldsymbol{s} \end{bmatrix} = \begin{bmatrix} \sin(\boldsymbol{\psi}_t^h) & \cos(\boldsymbol{\psi}_t^h) \\ \cos(\boldsymbol{\psi}_t^h) & -\sin(\boldsymbol{\psi}_t^h) \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_{c,t}^H \\ \boldsymbol{y}_{c,t}^H \end{bmatrix},$$

so

$$\begin{bmatrix} \boldsymbol{x}_{c,t}^H \\ \boldsymbol{y}_{c,t}^H \end{bmatrix} = \begin{bmatrix} \sin(\boldsymbol{\psi}_t^h) & \cos(\boldsymbol{\psi}_t^h) \\ \cos(\boldsymbol{\psi}_t^h) & -\sin(\boldsymbol{\psi}_t^h) \end{bmatrix} \begin{bmatrix} \boldsymbol{r} \\ \boldsymbol{s} \end{bmatrix} = \boldsymbol{r} \begin{bmatrix} \sin(\boldsymbol{\psi}_t^h) \\ \cos(\boldsymbol{\psi}_t^h) \end{bmatrix} + \boldsymbol{s} \begin{bmatrix} \cos(\boldsymbol{\psi}_t^h) \\ -\sin(\boldsymbol{\psi}_t^h) \end{bmatrix}.$$

This implies that

$$(\bar{C}_3 \cap \bar{C}_4) \cap \Omega = \Big\{ \omega \in \Omega \mid \begin{bmatrix} \boldsymbol{x}_{c,t}^H \\ \boldsymbol{y}_{c,t}^H \end{bmatrix} = \boldsymbol{r} \begin{bmatrix} \sin(\boldsymbol{\psi}_t^h) \\ \cos(\boldsymbol{\psi}_t^h) \end{bmatrix} + \boldsymbol{s} \begin{bmatrix} \cos(\boldsymbol{\psi}_t^h) \\ -\sin(\boldsymbol{\psi}_t^h) \end{bmatrix}, \\ |\boldsymbol{r}| \leq 0.5(L_t + \sqrt{(L_h)^2 + (W_h)^2}), |\boldsymbol{s}| \leq 0.5(W_t + \sqrt{(L_h)^2 + (W_h)^2}); \boldsymbol{\psi}_t^H \in [0, 2\pi) \Big\}$$

or, equivalently, that

$$(\bar{C}_3 \cap \bar{C}_4) \cap \Omega = \Big\{ \omega \in \Omega \mid (\boldsymbol{x}_{c,t}^H, \boldsymbol{y}_{c,t}^H) \in R_2(\boldsymbol{\psi}_t^H); \psi_t^H \in [0, 2\pi) \Big\}. \tag{12}$$

Replacing (11)-(12) into (10) yields

$$\begin{aligned} A &= \Big\{ \omega \in \Omega \mid (\boldsymbol{x}_{c,t}^H, \boldsymbol{y}_{c,t}^H) \in R_1 \cap R_2(\boldsymbol{\psi}_t^H); \psi_t^H \in [0, 2\pi) \Big\} \\ &= \Big\{ \omega \in \Omega \mid (\boldsymbol{x}_{c,t}^H, \boldsymbol{y}_{c,t}^H, \boldsymbol{\psi}_t^H) \in R_1 \cap R_2(\boldsymbol{\psi}_t^H) \times [0, 2\pi) \Big\}, \end{aligned}$$

which proves the Theorem. ∎

To end this section, we provide a numerical implementation of the results in Theorem 7.

A Numerical Implementation

The goal here is to compute $\mathrm{P}\{(\boldsymbol{x}_{c,t}^H, \boldsymbol{y}_{c,t}^H) \in R_1 \cap R_2(\boldsymbol{\psi}_t^H); \boldsymbol{\psi}_t^H \in [0, 2\pi)\}$. Formally,

$$\mathrm{P}\{(\boldsymbol{x}_{c,t}^H, \boldsymbol{y}_{c,t}^H) \in R_1 \cap R_2(\boldsymbol{\psi}_t^H); \boldsymbol{\psi}_t^H \in [0, 2\pi)\} \\ = \iiint 1_{\{(x,y,\psi) \in R_1 \cap R_2(\boldsymbol{\psi}_t^H) \times \in [0,2\pi)\}} f_{x_{c,t}, y_{c,t}, \psi_t} \mathrm{d}x \mathrm{d}y \mathrm{d}\psi.$$

Under Assumption 1, the above expression can be further simplified as follows:

$$\mathrm{P}\{(\boldsymbol{x}_{c,t}^H, \boldsymbol{y}_{c,t}^H) \in R_1 \cap R_2(\boldsymbol{\psi}_t^H); \boldsymbol{\psi}_t^H \in [0, 2\pi)\} \\ = \int 1_{\{\psi \in [0,2\pi)\}} \left(\iint 1_{\{(x,y,\psi) \in R_1 \cap R_2(\boldsymbol{\psi}_t^H)\}} f_{x_{c,t}, y_{c,t}} \mathrm{d}x \mathrm{d}y \right) f_{\psi_t} \mathrm{d}\psi. \tag{13}$$

To further simplify the expression above, knowledge of the specific type of distributions for each random variable would be needed. In the absence of emperical information, the following assumption will be made:

Assumption 3. *$\boldsymbol{x}_{c,t}^H$, $\boldsymbol{y}_{c,t}^H$, and $\boldsymbol{\psi}_t^H$ are uniformly distributed random variables.*

Since the world model provides estimates of the mean, μ, and variance, σ^2, for each measured variable, the probability density functions associated with $\boldsymbol{x}_{c,t}^H$, $\boldsymbol{y}_{c,t}^H$, and $\boldsymbol{\psi}_t^H \boldsymbol{x}_{c,t}^H$, $\boldsymbol{y}_{c,t}^H$, and $\boldsymbol{\psi}_t^H$ are given by:

$$f_{x_{c,t}} = \frac{1}{\sqrt{12}\sigma_{x_{c,t}^H}} 1_{\{x \in [\mu_{x_{c,t}^H} - \sqrt{3}\sigma_{x_{c,t}^H}, \mu_{x_{c,t}^H} + \sqrt{3}\sigma_{x_{c,t}^H}]\}},$$
$$f_{y_{c,t}} = \frac{1}{\sqrt{12}\sigma_{y_{c,t}^H}} 1_{\{y \in [\mu_{y_{c,t}^H} - \sqrt{3}\sigma_{y_{c,t}^H}, \mu_{y_{c,t}^H} + \sqrt{3}\sigma_{y_{c,t}^H}]\}},$$

and

$$f_{\psi_t} = \frac{1}{\sqrt{12}\sigma_{\psi_t^H}} 1_{\{\psi \in [\mu_{\psi_t^H} - \sqrt{3}\sigma_{\psi_t^H}, \mu_{\psi_t^H} + \sqrt{3}\sigma_{\psi_t^H}]\}}.$$

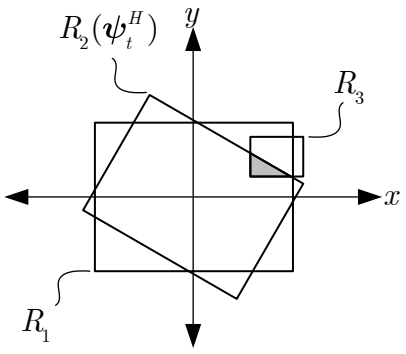

Figure 5: The gray area shows the region of integration associated with (14).

Replacing the above functions in (13) yields, after a few manipulations, the following:

$$\mathrm{P}\{(\boldsymbol{x}_{c,t}^H, \boldsymbol{y}_{c,t}^H) \in R_1 \cap R_2(\boldsymbol{\psi}_t^H)\boldsymbol{\psi}_t^H \in [0, 2\pi)\}$$
$$= \frac{1}{12\sqrt{12}\sigma_{x_{c,t}^H}\sigma_{y_{c,t}^H}\sigma_{\psi_t^H}} \int_{\mu_{\psi_t^H}-\sqrt{3}\sigma_{\psi_t^H}}^{\mu_{\psi_t^H}+\sqrt{3}\sigma_{\psi_t^H}} \left(\iint 1_{\{(x,y,\psi)\in R_1\cap R_2(\psi)\cap R_3\}} \mathrm{d}x\mathrm{d}y \right) \mathrm{d}\psi, \quad (14)$$

where $R_3 \subset \mathbb{R}^2$ is the rectangular subset of $\mathbb{R}^2$ given by:

$$R_3 \triangleq \{(x,y) \in \mathbb{R}^2 \mid x \in [\mu_{x_{c,t}^H} - \sqrt{3}\sigma_{x_{c,t}^H}, \mu_{x_{c,t}^H} + \sqrt{3}\sigma_{x_{c,t}^H}], y \in [\mu_{y_{c,t}^H} - \sqrt{3}\sigma_{y_{c,t}^H}, \mu_{y_{c,t}^H} + \sqrt{3}\sigma_{y_{c,t}^H}]\}.$$

Note that the expression between parenthesis in (14) is the area of the polygon created by intersecting R_1, $R_2(\psi)$, and R_3 (the intesection could also be empty). An example of this intersection is shown in Figure 5. Let $m(R_1 \cap R_2(\psi) \cap R_3)$ denote this area and observe that it is a function of the angle ψ. Although a formula could be derived to compute this area as a function of ψ, such an expression would not provide more insights or lead to simplifications. Hence, a numerical approach is better suited to compute (14).

To this end, let $1 \leq n \in \mathbb{Z}^+$ and set $\psi_i = \mu_{\psi_t^H} + \sqrt{3}\sigma_{\psi_t^H}\left(\frac{2i}{n} - 1\right)$, $i = 0, \ldots, n-1$. The right hand side of (14) can now be approximated as follows:

$$\mathrm{P}\{(\boldsymbol{x}_{c,t}^H, \boldsymbol{y}_{c,t}^H) \in R_1 \cap R_2(\boldsymbol{\psi}_t^H)\boldsymbol{\psi}_t^H \in [0, 2\pi)\} \approx \frac{1}{12n\sigma_{x_{c,t}^H}\sigma_{y_{c,t}^H}} \sum_{i=0}^{n-1} m(R_1 \cap R_2(\psi_i) \cap R_3) \quad (15)$$

where, for every ψ_i, $m(R_1 \cap R_2(\psi_i) \cap R_3)$ can be computed using standard functions for polygon intersection (see, e.g., Sutherland-Hodgman Polynomial Clipping Algorithm in [28]) and polygon area [29]. The calculation of (15) can be performed numerically by implementing Algorithm 1.

CONCLUSIONS

This article proposed a methodology to compute the "safety" of a vehicle quantitatively, so it can be used by automated vehicle for decision making and control. The methodology treats each interaction a vehicle has with other road user and road interactions as potential hazards and assigns each value of risk. The risk a hazard imposes on the vehicle in derived from the hazard's likelihood, hazardousness, and from the capability of the vehicle to avoid it. The article also offered a theoretical method to estimate the likelihood of two-dimensional collision hazards (an extension of the work in [9]) and recommendations on how to estimate hazard hazardousness.

Algorithm 1 InSTProbColEstimator: Instantaneous, Single Target, Host-Target Probability of Collision Estimator

Require: $W_h > 0, L_h > 0, W_t > 0, L_t > 0$ {host & target dimensions}

Require: $x^H_{c,t}, y^H_{c,t}, \sigma_{x^H_{c,t}} \geq 0, \sigma_{y^H_{c,t}} \geq 0$ {mean and variance of the target position w.r.t. the host}

Require: $\psi^H_t, \sigma_{\psi^H_t} \geq 0$ {mean and variance of the target's heading w.r.t. the host's vertical axis}

Require: $n \geq 1$ {number of discretization points for the range of $\boldsymbol{\psi}^H_t$ }

Ensure: $PCol =$ Probability of Collision

$PCol \leftarrow 0$ {$PCol$ is zero by default}

$$R_1 \leftarrow \begin{bmatrix} -0.5(L_h+\sqrt{(L_t)^2+(W_t)^2}) & -0.5(W_h+\sqrt{(L_t)^2+(W_t)^2}) \\ 0.5(L_h+\sqrt{(L_t)^2+(W_t)^2}) & -0.5(W_h+\sqrt{(L_t)^2+(W_t)^2}) \\ 0.5(L_h+\sqrt{(L_t)^2+(W_t)^2}) & 0.5(W_h+\sqrt{(L_t)^2+(W_t)^2}) \\ -0.5(L_h+\sqrt{(L_t)^2+(W_t)^2}) & 0.5(W_h+\sqrt{(L_t)^2+(W_t)^2}) \end{bmatrix} \{R_1 \text{ vertices, counterclockwise (CCW)}\}$$

$$R_3 \leftarrow \begin{bmatrix} x^H_{c,t}-\sqrt{3}\sigma_{x^H_{c,t}} & y^H_{c,t}-\sqrt{3}\sigma_{y^H_{c,t}} \\ x^H_{c,t}+\sqrt{3}\sigma_{x^H_{c,t}} & y^H_{c,t}-\sqrt{3}\sigma_{y^H_{c,t}} \\ x^H_{c,t}+\sqrt{3}\sigma_{x^H_{c,t}} & y^H_{c,t}+\sqrt{3}\sigma_{y^H_{c,t}} \\ x^H_{c,t}-\sqrt{3}\sigma_{x^H_{c,t}} & y^H_{c,t}+\sqrt{3}\sigma_{y^H_{c,t}} \end{bmatrix} \{R_3 \text{ vertices, CCW}\}$$

$[RisEmpty, R] \leftarrow PolyIntersect(R_1, R_3)$ {R: $R_1 \cap R_3$ CCW vertices. If empty, $RisEmpty = 1$}

if $\neg RisEmpty$ **then**

  for $i = 0$ to $n-1$ **do**

    $\psi \leftarrow \psi^H_t + \sqrt{3}\sigma_{\psi^H_t}\left(\frac{2i}{n} - 1\right)$

$$R_2 \leftarrow \begin{bmatrix} -0.5(L_t+\sqrt{(L_h)^2+(W_h)^2}) & 0.5(W_t+\sqrt{(L_h)^2+(W_h)^2}) \\ -0.5(L_t+\sqrt{(L_h)^2+(W_h)^2}) & -0.5(W_t+\sqrt{(L_h)^2+(W_h)^2}) \\ 0.5(L_t+\sqrt{(L_h)^2+(W_h)^2}) & -0.5(W_t+\sqrt{(L_h)^2+(W_h)^2}) \\ 0.5(L_t+\sqrt{(L_h)^2+(W_h)^2}) & 0.5(W_t+\sqrt{(L_h)^2+(W_h)^2}) \end{bmatrix} \begin{bmatrix} \sin(\psi) & \cos(\psi) \\ -\cos(\psi) & \sin(\psi) \end{bmatrix} \{R_2(\psi_i) \text{ CCW vertices}\}$$

    $[R_aEmpty, R_a] \leftarrow PolyIntersect(R, R_2)$ {R_a:$R_1 \cap R_3 \cap R_2(\psi)$ CCW vertices. If empty, $R_aEmpty = 1$}

    if $\neg R_aEmpty$ **then**

      $PCol \leftarrow PCol + PolyArea(R_a)$

    end if

  end for

  $PCol \leftarrow PCol/(12n\sigma_{x^H_{c,t}}\sigma_{y^H_{c,t}})$

end if

Further work is required to link the likelihood estimator with motion predictor models, and to validated the complete methodology first by detailed simulations and they by experimentation.

Acknowledgements
The authors would like to thank Stefanie Hair-Buijssen (TNO), Richard van der Horst, Aliaksei Laureshyn (Lund University), Erwin de Gelder (TNO), Jan-Pieter Paardekooper (TNO), Ron Snijders (TNO), and Jeroen Manders (TNO) for important information and discussions that sharpened the ideas in this article.

References

[1] NHTSA. (2017) 2016 fatal motor vehicle crashes: Overview. Accessed on March 07, 2019. [Online]. Available: https://www.nhtsa.gov/press-releases/usdot-releases-2016-fatal-traffic-crash-data

[2] S. Behere and M. Törngren, "A functional reference architecture for autonomous driving," *Information and Software Technology*, vol. 73, pp. 136 – 150, 2016.

[3] Y. Luo, A. K. Saberi, T. Bijlsma, J. J. Lukkien, and M. van den Brand, "An architecture pattern for safety critical automated driving applications: Design and analysis," in *2017 Annual IEEE International Systems Conference (SysCon)*, April 2017, pp. 1–7.

[4] A. Svensson and C. Hydén, "Estimating the severity of safety related behavior," *Accid Anal Prev*, vol. 38, no. 2, pp. 379 – 385, 2006.

[5] L. Zheng, K. Ismail, and X. Meng, "Traffic conflict techniques for road safety analysis: open questions and some insights," *Canadian Journal of Civil Engineering*, vol. 41, no. 7, pp. 633–641, 2014.

[6] F. Diederichs, T. Schüttke, and D. Spath, "Driver intention algorithm for pedestrian protection and automated emergency braking systems," in *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, Sep. 2015, pp. 1049–1054.

[7] E. Dagan, O. Mano, G. P. Stein, and A. Shashua, "Forward collision warning with a single camera," in *IEEE Intelligent Vehicles Symposium, 2004*, June 2004, pp. 37–42.

[8] ISO 26262-1:2018, road vehicles – functional safety – part 1: Vocabulary. ISO. [Online]. Available: https://www.iso.org/standard/68383.html

[9] T. van den Broek and J. Ploeg, "Collision warning system based on probability density functions,," in *WIT 2010 - 7th International Workshop on Intelligent Transportation*, March 2010, pp. 141–148.

[10] E. van Nunen, T. van den Broek, M. Kwakkernaat, and D. Kotiadis, "Implementation of probabilistic risk estimation for VRU safety," in *WIT 2011 - 8th International Workshop on Intelligent Transportation*, March 2011, pp. 149–155.

[11] D. Lord and S. Washington, *Chapter 1. Introduction*, 2018, ch. Chapter 1, pp. 1–10.

[12] M. Belin, R. Johansson, J. Lindberg, and C. Tingvall, "The vision zero and its consequences1," in *4th International Conference on Safety and the Environment in the 21st Century*, Nov. 1997, pp. 1–14.

[13] W. Wijnen, W. Weijermars, W. Vanden Berghe, A. Schoeters, R. Bauer, L. Carnis, R. Elvik, A. Theofilatos, A. Filtness, S. Reed, C. Perez, and H. Martensen, "Crash cost estimates for european countries," Eindhoven University of Technology, Deliverable 3.2 of the H2020 project SafetyCube, 2017, http://www.safetycube-project.eu/publications/.

[14] A. Laureshyn, T. De Ceunynck, C. Karlsson, A. Svensson, and S. Daniels, "In search of the severity dimension of traffic events: Extended Delta-V as a traffic conflict indicator," *Accid Anal Prev*, vol. 98, pp. 46–56, Jan 2017.

[15] A. P. Tarko, *Chapter 17. Surrogate Measures of Safety*, 2018, ch. Chapter 17, pp. 383–405.

[16] A. Laureshyn, A. Svensson, and C. Hydén, "Evaluation of traffic safety, based on micro-level behavioural data: Theoretical framework and first implementation," *Accident Analysis & Prevention*, vol. 42, no. 6, pp. 1637 – 1646, 2010.

[17] O. op den Camp, S. van Montfort, J. Uittenbogaard, and J. Welten, "Cyclist target and test setup for evaluation of cyclist-autonomous emergency braking," *International Journal of Automotive Technology*, vol. 18, no. 6, pp. 1085–1097, Dec 2017.

[18] S. S. Mahmud, L. Ferreira, M. S. Hoque, and A. Tavassoli, "Application of proximal surrogate indicators for safety evaluation: A review of recent developments and research needs," *IATSS Research*, vol. 41, no. 4, pp. 153 – 163, 2017.

[19] A. Laureshyn, "Application of automated video analysis to road user behaviour," Ph.D. dissertation, Lund University, 2010.

[20] F. A. Mullakkal-Babu, M. Wang, H. Farah, B. van Arem, and R. Happee, "Comparative assessment of safety indicators for vehicle trajectories on highways," *Transportation Research Record*, vol. 2659, no. 1, pp. 127–136, 2017.

[21] K. Ismail, T. Sayed, and N. Saunier, "Methodologies for aggregating indicators of traffic conflict," *Transportation Research Record*, vol. 2237, no. 1, pp. 10–19, 2011.

[22] S. Yoshida, T. Hasegawa, S. Tominaga, and T. Nishimoto, "Development of injury prediction models for advanced automatic collision notification based on japanese accident data," *International Journal of Crashworthiness*, vol. 21, no. 2, pp. 112–119, 2016.

[23] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge: Cambridge University Press, 2009.

[24] V. L. Klee, Jr, "Strict separation of convex sets," *Proc. American Mathematical Society*, vol. 7, no. 4, pp. 735–737, August 1956.

[25] A. A. Ahmadi. (2016) Convex and conic optimization, lecture notes 5. Accessed on March, 2018. [Online]. Available: https://goo.gl/d4qTCY

[26] S. Gottschalk, "Separating axis theorem," Department of Computer Science, UNC Chappel Hill, Tech. Rep. TR96-024, 1996.

[27] P. J. Schneider and D. H. Eberly, *Geometric Tools for Computer Graphics*. Boston: Elsevier Science, 2003.

[28] J. D. Foley, A. van Dam, S. K. Feiner, and J. F. Hughes, *Computer Graphics: Principles and Practice*. Addison-Wesley Publishing Company, 1996.

[29] P. Bourke. (1988) Calculating the area and centroid of a polygon. Accessed on April 26, 2018. [Online]. Available: http://paulbourke.net/geometry/polygonmesh