Towards threat metric evaluation in complex urban scenarios*

Patrick Schneider[1] and Martin Butz[2] and Christian Heinzemann[2] and Jens Oehlerking[2] and Matthias Woehrle[2]

***Abstract*— Threat metrics are used to measure the criticality of traffic situations. Such metrics rely on models predicting future behavior of other traffic participants. These models need to be calculated with respect to lanes for handling oncoming traffic, but in urban scenarios, lanes are often not explicitly defined. In this paper, we present a first modeling and conformance testing approach for criticality metrics in urban scenarios. In addition, we analyze the influence of lane association on the quality of the models used for threat prediction. We evaluate our approach based on an urban intersection from the inD dataset. Our key finding is that such metrics are in general prone to both false positives and negatives if maps with static lanes are used.**

I. Introduction

Urban automated driving requires automated vehicles (AV) to master a complex environment with many different kinds of traffic participants like other cars, pedestrians, cyclists etc. For all traffic situations that the system may encounter, we expect the AV to not cause dangerous situations or even accidents. To this end, various threat metrics have been proposed in literature [1] to measure how critical a certain situation at hand is. Such criticality estimates provided by threat metrics provide valuable input for scenario-based testing activities and situational risk-aware driving [2].

Relying on threat metrics for assessing the behavior of the AV, however, is only meaningful if the metrics provide meaningful criticality estimates. To this end, all threat metrics rely on predictive models for the other traffic participants for computing the criticality estimates [1]. As a consequence, the reliability and suitability of the calculated criticality estimate significantly depends on the quality of the underlying predictive models and their ability to make a reasonable prediction of the future behavior of the other traffic participants. While this works reasonably well in a highway setting [3], these results do not directly transfer to more complex, structured environments like urban intersections. This is because in such scenarios, we often do not have explicit lanes and we frequently encounter traffic participants that operate non-lane-based like pedestrians and cyclists.

Despite a lot of research regarding threat metrics [1], [4], [5], [6], [7], [8] and a process for criticality analysis in urban settings [2], the assessment of suitable models for threat metrics is still not sufficiently solved. Approaches using motion models and corresponding conformance testing techniques [9] focused on free-space behavior of pedestrians [10] or vehicles [11] neglecting the influence of lane-association on conformance.

In this paper, we present and evaluate an approach for systematic model definition and conformance testing for an urban intersection with the goal of using the derived models in threat metrics. As our main contribution, we illustrate how model definition and model conformance testing can be applied to urban settings where lane-based behavior is demanded, e.g., for being able to handle oncoming traffic. A key difficulty, however, is that lanes are often not explicitly marked in these settings and emerge from the interaction of different vehicles, e.g., when turning at an intersection. Here, we particularly evaluate the impact of lane association on conformance and discuss the resulting influence on threat prediction at urban intersections.

We evaluate our approach on an urban intersection from the inD dataset [12] shown in Fig. 1, located in Aachen, Germany. For the model definition, we use all available vehicle data for this intersection. For the evaluation of the effect of lane association on conformance, we focus on situations where a vehicle needs to let pass oncoming traffic before making a left turn. The key take-aways from our evaluation are: (1) the (virtual) lanes that other vehicles use depend on the concrete situation such that a static map is insufficient and (2) threat prediction is highly influenced by meaningful derivation of (virtual) lanes for traffic participants, particularly for turn maneuvers.

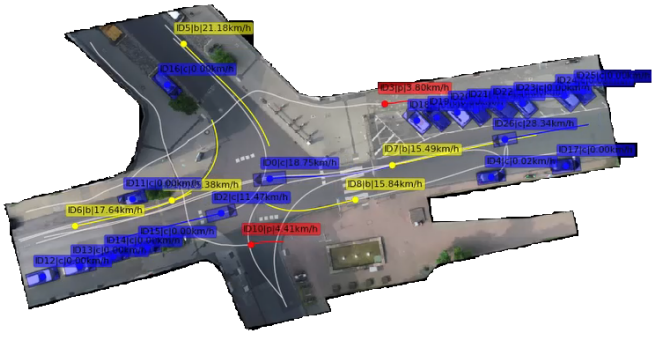

Fig. 1. Example scene on the used intersection from the inD dataset. The image is taken from the background images supplied together with the inD dataset.

The remainder of this paper is structured as follows: Section II introduces background and related works before we describe our modeling approach in Section III. Thereafter, we present the results of our evaluation in Section IV before concluding the paper in Section V.

*The research leading to these results is funded by the German Federal Ministry for Economic Affairs and Energy within the project 'VVM - Verification & Validation Methods for Automated Vehicles Level 4 and 5'.

[1]First and corresponding author. University of Stuttgart, Germany. Thesis work at Robert Bosch GmbH. st160023@stud.uni-stuttgart.de

[2]Robert Bosch GmbH, Corporate Research, 71272 Renningen, Germany firstname.lastname@de.bosch.com

II. Background and Related work

In the following, we discuss background and related work from two areas: (1) criticality analysis for urban scenarios and (2) models and model conformance for lane-based traffic.

As laid out above, the vast complexity of the urban traffic system poses a challenge for the development and validation of autonomous systems and has been a focus of research over the last decades. Dahl et al. [1] present a deep survey on the use of threat metrics for the assessment of traffic situations. Huber et al. [7] use a multi-dimensional assessment approach for the evaluation of traffic situations with the aim to select scenarios for testing. Many approaches and studies striving to give a framework for the safe operation of autonomous vehicles have been presented like for example RSS [13] and SFF [14]. To enable scenario-based testing, as proposed by the PEGASUS project[1], the structuring of the scenario-space is a key challenge. Hallerbach et al. [6] present a simulation-based approach for the identification of the critical scenarios. Jesenski et al. [8] use causality groups to reduce the complexity of the traffic system to enable the scalable generation of statistical evidence. With the aim of deriving a suitable arguable structure the criticality analysis proposed in the VVM project[2] by Neurohr et al. [2] focuses on the identification of relevant phenomena by means of threat metrics and their causal chains contributing to an increase of criticality in a traffic situation.

Threat metrics rely on some form of model for inferring the future behavior of other traffic participants for calculating a criticality for the traffic situation at hand. These models naturally affect the results of the evaluation. Junietz et al. [4] discuss such effects for a single-track model for vehicles. This effect can be interpreted as a form of uncertainty. The effects of uncertainties within threat measures for driver assistance have been studied by Stellet et al [15].

The task of checking whether the predictions of such models actually represent the behavior of traffic participants in the real-world is called conformance testing [9]. Often, the models used in threat metrics assume one particular behavior like applying a constant velocity or constant acceleration. Obviously, such models do not conform to all possible real world behaviors. As a solution, set-based models based on differential inclusions can be used that are amenable to reachability analysis [16]. Such models take intervals of possible positions and velocities as inputs (e.g., if these cannot be measured fully accurately), consider a bounded set of possible actions (e.g., accelerations) reflecting the uncertainty of the future behavior, and yield a set of possible future locations and velocities for a traffic participant. Koschi and Althoff have extended such models towards lane-based behavior [17] where the future locations of traffic participants are projected along their lanes potentially considering also lane changes. Corresponding extensions of threat metrics and RSS to set-based models also exist [18], [19]. While conformance for set-based models has been defined [11], model conformance has not yet been considered for the lane-based case. Hartung et al. [20] analyze models of free-space driving maneuvers and show conformance to measurements from a test vehicle. Liu et al. [10] introduce a free-space pedestrian model and show conformance based on a dataset. Koschi et al. [21] consider pedestrians in an urban environment, but do not incorporate lanes or other infrastructure in their conformance test. Zecher et al. [22] validate bounds on lateral movement on highways based on a large dataset, but without consideration of lanes.

[1] https://www.pegasusprojekt.de
[2] https://www.vvm-projekt.de/

III. Modeling Framework

In this section, we describe our modeling framework, consisting of a model of the map and motion models for the traffic participants. Both are designed with reachability analysis in mind, i.e., the motion models are based on differential inclusions [16]. Furthermore, the motion models for vehicles are described in road coordinates with longitudinal and lateral axes instead of global coordinates, to make them usable independent of the precise map. This in turn means that the map needs to provide a road coordinate system for every road in the scenario, along which the models can be interpreted.

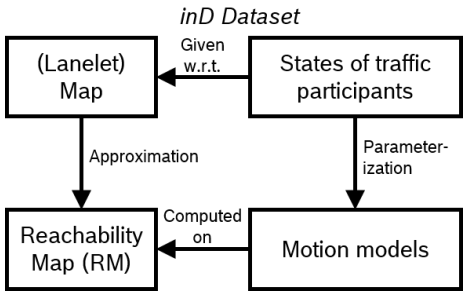

Fig. 2. Overview of modeling.

Figure 2 shows our general process for the modeling. Due to the absence of a road coordinate system, we cannot use the lanelet-based map as provided with the dataset directly. Therefore, we first (manually) have to create a map that approximates these lanelets and can be used for the reachability computation. Since threat metrics always rate the traffic situation from the viewpoint of one vehicle while the dataset is viewpoint agnostic, we need to explicitly choose one vehicle from an interaction relative to which the criticality is computed. We refer to this vehicle as the *ego* vehicle in the following and state how it was selected where necessary.

In the following, we describe the map (III-A), the motion models (III-B), and how they are used in reachability analysis (III-C) and for metric calculation (III-D). We conclude the section by describing the parameter identification (III-E) for the motion models based on the data from the inD dataset.

A. Map

As discussed above, the map that is needed for the reachability computation differs from the lanelet-based map

that is provided with the dataset. Figure 3 shows the original model in orange, superimposed with our map model in red (driving lanes) and purple (pedestrian crossings). Instead of polygons, our map is based on road segments that are either straight or have a fixed curvature, and which are sequentially composed into a crossroads structure. Each such segment has a fixed width, and is also further subdivided into lanes of fixed width, which are not visualized in Fig. 3. As such each segment naturally provides a road coordinate system with a lateral and a longitudinal axis, with lateral zero typically being the road's center line. The connections between different road segments are represented by predecessor/successor relations. This modeling approach is compatible to the one taken e.g. in the OpenDrive standard [23]. Different connections within the crossroads are also modeled as separate roads, as can be seen in the figure.

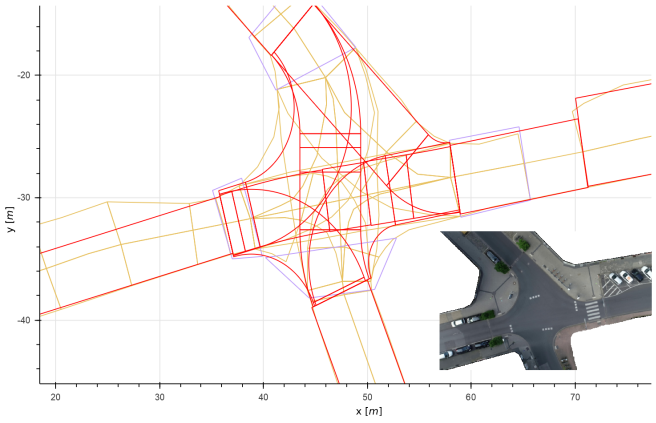

Fig. 3. Reachability map (RM) in red, lanelet-based map in yellow and pedestrian crossings in purple.

This modeling paradigm has several consequences. Firstly, we need to approximate the actual road geometry somewhat more loosely than with a lanelet-based approach. Secondly, the choice of road connections inside the crossroads (e.g. for turning left) already models an assumed implicit lane that will be used for prediction of vehicle behavior. However, the advantage of this modeling approach lies in the fact that we express movement models for vehicles purely in terms of longitudinal and lateral velocities and accelerations. This establishes a separation of two modeling decisions: which lane sequence a vehicle is predicted to follow (which is encoded in the map), and with which dynamics within the lane sequence it is then predicted to travel (which is encoded in the motion model). Reachability computations are then only done for the motion model in road coordinates, which greatly simplifies the reachability computations, essentially turning them into a two-dimensional generalization of approaches such as RSS [13]. Reachable sets computed based on differential inclusion dynamics, and then intersected with the geometric shape of the lane sequence for the vehicle to ensure the prediction stays within lane bounds. Since the predictive motion models we use are the same regardless of the actual location of the vehicle on the road network, parameterization for vehicle models can be done based on the entire data set.

B. Motion Models

Models for traffic participants need to be divided into two groups: those that follow lane structures and those that do not. The first class would contain subjects such as cars, trucks, motorbikes or bicycles when following a lane, while the second class would for instance include pedestrians crossing the road. The assignment of a traffic participant to one of these classes is in itself a modeling decision, as e.g. a pedestrian on a sidewalk could also be described using a lane-based model. In the following, we restrict ourselves to models of lane-based traffic participants and their association to lanes.

1) Set-based models: As discussed above, the models for vehicles are described in road coordinates. In the following, we denote the longitudinal position by s and the lateral position by t. Velocities and accelerations are denoted by v_s, v_t, a_s, and a_t, respectively.

Models are described by *differential inclusions* using these quantities as state variables. Differential inclusions can essentially be understood as differential equations imbued with set-based uncertainty in the dynamics to reflect the fact that we do not know precisely how a vehicle is going to move. This means that all model states are understood to be set-valued. We conservatively model the longitudinal and lateral dynamics as uncoupled differential inclusions, i.e., they are completely separate. In the following, τ denoted the (future) point in time, for which the predictive model is evaluated.

For the longitudinal dynamics of the ego vehicle, we define three different models that we will use in our evaluation. The first one simply has constant velocity.

$$\dot{s}(\tau) \in [v_s^L, v_s^U] \tag{1}$$

The second one assumes constant acceleration instead.

$$\dot{s}(\tau) = v_s(\tau), \dot{v}_s(\tau) \in [a_s^L, a_s^U] \tag{2}$$

The last model has a time period of constant velocity (the reaction time τ_0), followed by constant (negative) acceleration modeling braking.

$$\begin{aligned} \dot{s}(\tau) &\in [v_s^L, v_s^U] \text{ if } \tau <= \tau_0 \\ \dot{s}(\tau) = v_s(\tau), \dot{v}_s(\tau) &\in [a_s^L, a_s^U] \text{ otherwise.} \end{aligned} \tag{3}$$

Note that we use the term *constant* to describe that the *interval* in which velocity or acceleration can lie in the prediction are constant. The value in the interval can change at every point in time τ. For the experiments that follow we actually chose point intervals for the ego vehicle, assuming the ego knows its states and follows its intentions exactly. Non-point intervals could be used account for uncertainties, e.g. of the controller. Using non-point intervals makes no technical difference for the reachability computation in our framework.

For the lateral dynamics of the ego vehicle, we use a model with constant velocity:

$$\dot{t}(\tau) \in [v_t^L, v_t^U] \tag{4}$$

For other vehicles, we always use a model with constant longitudinal acceleration according to Eq. (2). In this case, the acceleration interval is not a point interval and accounts for the possible braking and accelerating maneuvers of the vehicle. The lateral model for other vehicles is again given by Eq. (4).

2) Lane association: The reachable set computations from the models above need to be intersected with the lane sequences we expect a vehicle to follow. Without performing this intersection, the models would always predict possible collisions manifesting in unreasonably high criticality in case of oncoming traffic. Thus, we need to establish lane sequences for all vehicles.

Firstly, this means that it is required to associate points from the data set with road segments of our map. In order to do this, we match the vehicle to the road segment on which its reference point is located. For this, we only consider road segments that correspond to the maneuver being examined, e.g., only using the sequence of road segments running from east to west for the ego vehicle. Secondly, we need to establish all the road sequences originating in this initial road segment which are to be considered for the prediction. This is done by enumerating all paths in the graph structure given by the map's successor relations. A separate prediction is done for each such sequence. Thirdly, admissible lanes within the road segments are determined as follows: if a vehicle is located in the opposite lane within its initial road segment, the model permits the vehicle to use both lanes, otherwise it will only use the lane running in the "correct" direction. This means that our models is conservative in case a vehicle is not staying within the modeled lane bounds, by allowing the model to use the entire opposite lane as well. Less conservative models (e.g., just widening the lane enough to fit the vehicle) are possible, but were not implemented in the course of this work.

C. Reachability Analysis

Reachability analysis is a technique to compute reachable system states for (among others) motion models given as differential inclusions as described in Section III-B. Instead of sampling-based prediction symbolic representations of sets are used and predicted, resulting in predictions that are guaranteed to include all states that are mathematically reachable. In mathematical terms, given a differential inclusion of the form

$$\dot{x}(\tau) \in F(x(\tau)),$$

the result for each point in time τ is a set $R(\tau)$ that includes all states $x(\tau)$ that are possible under these dynamics. There exist various computation methods for reachable sets, i.e., using zonotopes [24] or support functions [25] as set representations. Because our models have very simple dynamics, we instead simply bound each state variable (position and velocity) by an interval, so for a vehicle state $x = (s, t, v_s, v_t)$ we have

$$R(\tau) = [s^L, s^U] \times [t^L, t^U] \times [v_s^L, v_s^U] \times [v_t^L, v_t^U]$$

where the upper and lower bounds for each time τ can be computed by using interval arithmetic. We also only compute the reachable set for equidistant discrete time steps, every 0.04 seconds, which is also the distance between the data samples in the inD dataset. In addition, we restrict the reachable sets to fit the lanes in each step. Here, we use all lanes the vehicle is currently associated to as described in Section III-B.2. Since each road has a fixed width, this simply amounts to restricting the lateral position interval $[t^L, t^U]$ to the lane width. In case a reachable state can lie on two roads of different widths, we conservatively choose the wider road.

D. Metrics

In the following, we use three computation principles for metrics which are built upon the differential inclusion models. Let R_{ego} be a reachable set for the ego vehicle and let R_v be a reachable set for another vehicle. For a reachable set R, let $\bar{R}$ denote the coordinate transformation of its position component onto the global x-y-plane (i.e., the set of reachable positions at a point in time, disregarding velocities). For other traffic participants, we always use the vehicle model from Eqs. (2) and (4).The first metric is *time-to-collision* like and computes the earliest time

$$M_{TTC} = \min\{\tau \mid \bar{R}_{ego}(\tau) \cap \bar{R}_v(\tau) \neq \emptyset\} \tag{5}$$

at which the reachable sets of two traffic participants intersect. For the computation we use the constant velocity model from Eq. (1) and the lateral model from Eq. (4) for the ego vehicle. The second metric is similar in structure to the *brake threat number* [1]. It quantifies the deceleration that is needed to avoid a collision, using Eq. (2) for the longitudinal and Eq. (4) for the lateral behavior. Let R_{a_s} be the reachable set computed for a given value of $a_s = a_s^L = a_s^U$. Thus we set

$$M_{BTN} = \max\{a_s/a_{br} \mid \forall \tau : \bar{R}_{a_s}(\tau) \cap \bar{R}_v(\tau) = \emptyset\}, \tag{6}$$

where a_{br} is the parameter describing the maximal deceleration rate possible. The third metric models the *time-to-react* as suggested in [26]. The model for the ego vehicle is given by Eqs. (3) and (4). For a given τ_0, let R_{τ_0} be the computed reachable set when setting the constant deceleration parameter $\dot{v}_s(\tau)$ in Eq. (3) accordingly. We thus define

$$M_{TTR} = \max\{\tau_0 \mid \forall \tau : \bar{R}_{\tau_0}(\tau) \cap \bar{R}_v(\tau) = \emptyset\}. \tag{7}$$

This metric models the last point in time at which the ego vehicle has to start braking, while still being able to avoid a collision. These metrics naturally depend on the parameterization of the underlying models. The parameterization for all bound parameters as part of the metric is described in the following.

E. Parameterization

The defined metrics and the underlying models require a proper parameterization for their evaluation. Depending, on which model is used, the bounds for the acceleration a_s^L, a_s^U or velocity v_t^L, v_t^U, v_s^L, v_s^U are required. Table I lists the various models for the ego as well the other traffic participants and shows the parameters used.

TABLE I

PARAMETERIZATION FOR **EGO** MODELS USED IN THE METRICS M_{TTC}, M_{BTN} AND M_{TTR} AS WELL AS **OTHER** VEHICLES. LEGEND: *VALUE TAKEN FROM IND DATA AT TIME t (BOTH VALUES EQUAL), **COMPUTED BY METRIC (BOTH VALUES EQUAL)

Parameter	M_{TTC} ego	M_{BTN} ego	M_{TTR} ego	others
τ_0	–	–	computed	–
v_s^L, v_s^U	inD data*	–	inD data*	–
a_s^L, a_s^U	–	computed**	-7, -7	-2.04, 1.88
v_t^L, v_t^U	-0.51, 0.49	-0.51, 0.49	-0.51, 0.49	-0.51, 0.49
a_{br}	–	-7	–	–

We gather all data from vehicles in all intersections contained in the inD dataset. For practical purposes, the data presented here is restricted to cars as there is much more data available. Table II summarizes the distribution using quantiles.

TABLE II

QUANTILES FOR VELOCITY AND ACCELERATION USED BY MODEL.

class	quantile	lateral velocity	longitudinal acceleration
car	Q_{02}	-0.51	-2.04
	Q_{50}	0.02	0.02
	Q_{98}	0.49	1.88

We focus on lateral velocity and longitudinal acceleration, as these need to be fitted from data. We determine upper and lower bounds and exclude long tails that partially (*i*) either include implausible values, *e.g.* due to measurement or preprocessing errors or (*ii*) excessively reckless behavior that we want to exclude. In particular, we use the 2% quantile (Q_{02}) as a lower bound and the 98% quantile (Q_{98}) as an upper bound for the respective model parameters. By using the 2% quantile, we exclude hard braking events. Since we only look at scenarios with crossing traffic, not follow-the-lane, this should have no strong effect, but will be further explored in the future. The upper 98% quantile is quite permissive on the behavior of other vehicles and one could further investigate effects of restrictions on maximal velocity.

For the switched model, there is an initial velocity that we use from the dataset plus we need to consider bounds.

IV. CONFORMANCE TESTING AND THREAT METRICS ON AN IND INTERSECTION

In this section, we describe the testing procedures to ascertain the quality of the modeling and parameterization, and the actual application of the models for the computation of threat metrics.

While all the data from all 4 intersections included in the inD dataset was used for parametrization, we perform the tests and the computation of the metrics on a single intersection, concretely intersection 2 as depicted in Figure 3. This intersection is the most suitable one for our experiments, since it features the highest amount of vulnerable road users both absolute (3799) as well as normalized per time (15.63 min^-1). Data is not filtered *i.e.* raw values from inD are used.

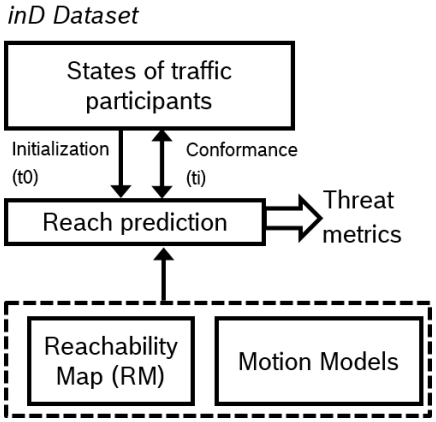

Fig. 4. Overview of application of framework.

A. Conformance Test

As shown in Figure 4, conformance is tested between the reach predictions and the actual states of traffic participants. Reach predictions are based on reachable sets using motion models with the parameter bounds computed above and the map, especially lane information, as presented above. For each conformance test, models are initialized with traffic participant states (at t_0) taken from the dataset. Starting from t_0, we compute reachable sets for all models for discrete steps of time $\delta T := 400ms$ (or 10 frames respectively) into the future. We select a horizon of $T_h := 2s$ resulting in a number of steps of $n = T_h/\delta T := 5$. Note that we select a large step size of 10 frames, to reduce the number of necessary computations, since each results in a dedicated reachability computation and therefore computational burden. We leave a more detailed exploration w.r.t. temporal resolution for future work.

We perform two different conformance checks: (*i*) we perform a basic model parameter conformance check to see whether longitudinal speed is contained in our acceleration-based model and (*ii*) a complementary lane-based check to see whether vehicles are located on a corresponding lane, based on the lane association as discussed in III-B.2.

1) Conformance test of vehicle dynamics: For this intersection, our conformance test includes 2184 vehicles[3] with a total number of about 8 million relevant frames. We perform a basic model parameter conformance check as discussed above. We can see that longitudinal speed does not conform to the acceleration-based model in 39274 out of 789316 frames and correspondingly for 866 out of 2184 cars. As we see, using the quantiles Q_{02} and Q_{98} of the vehicle's longitudinal acceleration for model parameterization results

[3]excluding parked vehicles

in about 5% failed checks where the actual velocity is not in the predicted range. Note that these are not due to individual vehicles, but rather distributed across a large number of cars.

2) Conformance test of lane association: Since the metrics are defined on pairs of vehicles, we always perform a pairwise selection of relevant traffic participants. For the considered scenario we perform the following steps. First, we remove parked cars. Second, we identify crossing vehicles, starting on west road and ending on north road, and ego vehicles, starting on east road and ending on west road. Third, we join these crossing vehicles with ego vehicles if we can identify that they meet in the actual intersection area defined by the map. Based on this approach we select 23 scenarios for further analysis. Within these, there are 22 and 21 unique tracks for the ego and the other vehicle, respectively. For both sets of tracks, we perform a lane association conformance check, see Sec. III-B.2. This means for each frame we take the predicted lanes as discussed above and check whether the next 5 frames are included in these lanes. For our analysis, we differentiate an inclusion check considering the vehicle geometry and one just considering the reference point. This allows us to study the difference between fails due to prediction and fails due to geometry.

Figure 5 summarizes results for other vehicles (top) and ego vehicles (bottom) with histograms of conformance check success rates with and without considering the shape. We can see for both histograms that considering shape has a large impact. When just considering the reference point, conformance is close to perfect. When considering the actual size of objects, the conformance success rates deteriorate considerable and most tracks suffer from failed conformance checks. Hence, one approach for increasing conformance is to increase lane width. This however impacts threat assessment, making it more conservative. A further analysis on the impact of lane width can be found in the following section. Additionally, for the other vehicle, which conducts a turning maneuver, there are considerable conformance check fails even without considering the shape. This suggests that better lane models are necessary.

B. Analysis of Crossing Scenarios

Table III shows the number of scenarios that are selected by individual threat metrics. For the scenario as described above the results are given in the bold line. The metrics identify the same 10 scenarios, M_{TTC} identifies one more, and a distance-based approach identifies three more.[4]

As a reference, we manually labeled each scenario on a 4-point scale: not critical (1), rather not critical (2), rather critical (3) and critical (4). Only one scenario is critical, four are rather critical and the rest are (rather) not critical.

Out of the 10 scenarios, 3 are (rather) critical. One additional rather critical scenario is flagged by M_{TTC} and the distance-based approach. The remaining rather critical scenario is never detected.

[4]Since the metrics identify the same scenarios, a further pairwise analysis did not further constrain the set of identified scenarios.

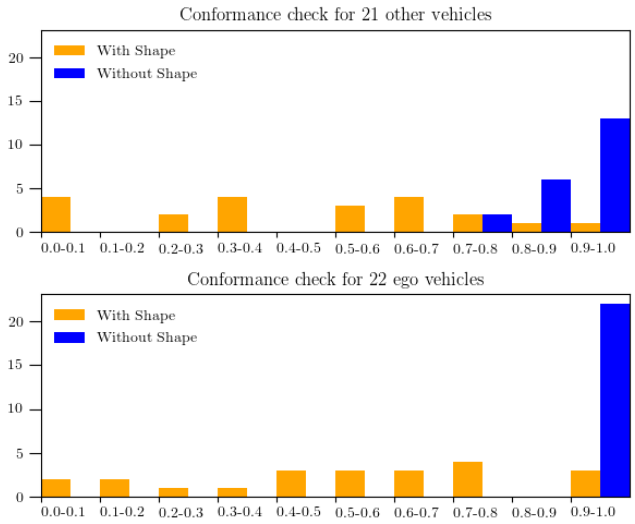

Fig. 5. Lane conformance success rate for other (top) and ego vehicles (bottom)

TABLE III

THRESHOLDS PER METRIC AND THE RESULTING NUMBER OF IDENTIFIED SCENARIOS (OUT OF 23 OVERALL SCENARIOS), WITH ORIGINAL LANE WIDTH AND OUTSIDE LANE BOUNDARIES MOVED BY 0.5 AND 1M TO THE OUTSIDE, RESPECTIVELY

	D_{min}	M_{TTC}	M_{TTR}	M_{BTN}
threshold	5m	5s	2s	0.3
identified scenarios (orig.)	**14**	**11**	**10**	**10**
identified scenarios (+0.5m)	17	12	12	12
identified scenarios (+1m)	17	12	12	12

All metrics are not particularly sensitive. Let us look at the example of M_{TTC}. It flags 11 scenarios for 4 to be identified, *i.e.* $\approx$ 36%.

So, given that our test is not particularly sensitive, how can it be that we do not arrive at perfect accuracy? We further analyze this particular scenario, which is actually the one shown in Figure 1. On further analysis, we determine the reason in the reachability approach used: Here reachable sets are projected onto lanes as modeled in Figure 3. However, on these intersections, crossing traffic participants do not necessarily follow these lanes. In fact, many vehicles rather continue to drive further along the straight line and then perform a sharper turn. The reachability analysis is myopic to this and projects in these cases onto the wrong lane. This weakens the analysis and leads to the missed detection.

Table III also shows the number of identified scenarios, when the lane boundaries are extended by 0.5m or 1m respectively. In these cases, the rather critical scenario that was originally only detected by M_{TTC} is now also detected by the other metrics. However, detection comes at the cost of additional reduction in sensitivity (all are now at 33%). Note that there is no difference between the cases 0.5m and 1m, indicating that vehicles were leaving the original lane only slightly.

As we can see a good trade-off between accuracy and

sensitivity is hard to achieve with static lane models. From this perspective, an approach based on dynamic lanes that considers driver's intention seems more plausible. For example, one can adapt the lane width depending on the position of the vehicles at the point in time for which we compute the metric. However, determining such dynamic lanes needs to be safe and reliable if threat analysis shall rely on this information.

V. Conclusion

In this work, we studied threat metrics for urban autonomous driving on the inD dataset. The computation of threat metrics is based on reachability computations. To this end, we presented modeling of actors and the map of an intersection. We showed how to parameterize these models and tested the resulting models for conformance. Our evaluation showed that the threat metrics allow us to identify most of the critical cases. However, they also identify many uncritical scenarios. A particular challenge that we have identified is the impact of lane models for the predictions, necessitating lane models that depend on the current state of the traffic participants. For all models that we have studied, there is a fundamental trade-off between accuracy (safety) and sensitivity (performance), requiring further study.

References

[1] J. Dahl, G. R. de Campos, C. Olsson, and J. Fredriksson, "Collision avoidance: A literature review on threat-assessment techniques," *IEEE Transactions on Intelligent Vehicles*, vol. 4, no. 1, pp. 101–113, March 2019.

[2] C. Neurohr, L. Westhofen, M. Butz, M. Bollmann, U. Eberle, and R. Galbas, "Criticality analysis for the verification and validation of automated vehicles," *IEEE Access*, vol. 9, pp. 18 016–18 041, 2021.

[3] P. Schneider, M. Butz, C. Heinzemann, J. Oehlerking, and M. Woehrle, "Scenario-based threat metric evaluation based on the highd dataset," in *2020 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2020.

[4] P. Junietz, F. Bonakdar, B. Klamann, and H. Winner, "Criticality metric for the safety validation of automated driving using model predictive trajectory optimization," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, Nov 2018, pp. 60–65.

[5] 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. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0386111217300286

[6] S. Hallerbach, Y. Xia, U. Eberle, and F. Koester, "Simulation-based identification of critical scenarios for cooperative and automated vehicles," apr 2018. [Online]. Available: https://doi.org/10.4271/2018-01-1066

[7] B. Huber, S. Herzog, C. Sippl, R. German, and A. Djanatliev, "Evaluation of virtual traffic situations for testing automated driving functions based on multidimensional criticality analysis," in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, 2020, pp. 1–7.

[8] S. Jesenski, N. Tiemann, J. E. Stellet, and J. M. Zöllner, "Scalable generation of statistical evidence for the safety of automated vehicles by the use of importance sampling," in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, 2020, pp. 1–8.

[9] H. Roehm, J. Oehlerking, M. Woehrle, and M. Althoff, "Model conformance for cyber-physical systems: A survey," *ACM Trans. Cyber-Phys. Syst.*, vol. 3, no. 3, Aug. 2019. [Online]. Available: https://doi.org/10.1145/3306157

[10] S. B. Liu, H. Roehm, C. Heinzemann, I. Lütkebohle, J. Oehlerking, and M. Althoff, "Provably safe motion of mobile robots in human environments," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems*, ser. IROS'17. IEEE Computer Society, Sept 2017, pp. 1351–1357.

[11] H. Roehm, J. Oehlerking, M. Woehrle, and M. Althoff, "Reachset conformance testing of hybrid automata," in *Proceedings of the 19th International Conference on Hybrid Systems: Computation and Control*, ser. HSCC '16. New York, NY, USA: ACM, 2016, pp. 277–286.

[12] J. Bock, R. Krajewski, T. Moers, S. Runde, L. Vater, and L. Eckstein, "The inD dataset: A drone dataset of naturalistic road user trajectories at german intersections," in *arXiv preprint arXiv:1911.07602*, 2019.

[13] S. Shalev-Shwartz, S. Shammah, and A. Shashua, "On a formal model of safe and scalable self-driving cars," 2018.

[14] D. Nistér, H.-L. Lee, N. Ng, and Y. Wang, *The Safety Force Field*, NVIDIA, Mar. 2019.

[15] J. E. Stellet, J. Schumacher, W. Branz, and J. M. Zöllner, "Uncertainty propagation in criticality measures for driver assistance," in *2015 IEEE Intelligent Vehicles Symposium (IV)*, 2015, pp. 1187–1194.

[16] M. Althoff and J. M. Dolan, "Online verification of automated road vehicles using reachability analysis," *IEEE Transactions on Robotics*, vol. 30, no. 4, pp. 903–918, Aug 2014.

[17] M. Koschi and M. Althoff, "Spot: A tool for set-based prediction of traffic participants," in *2017 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, June 2017, pp. 1686–1693. [Online]. Available: http://koschi.gitlab.io/spot/

[18] S. Söntges, M. Koschi, and M. Althoff, "Worst-case analysis of the time-to-react using reachable sets," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, 2018, pp. 1891–1897.

[19] P. F. Orzechowski, K. Li, and M. Lauer, "Towards responsibility-sensitive safety of automated vehicles with reachable set analysis," in *2019 IEEE International Conference on Connected Vehicles and Expo (ICCVE)*, 2019, pp. 1–6.

[20] M. Hartung, D. Hess, R. Lattarulo, J. Oehlerking, J. Pérez, and A. Rausch, *D5.2 – Report on Conformance Testing of Application Models*, June 2017. [Online]. Available: https://cps-vo.org/node/39012

[21] M. Koschi, C. Pek, M. Beikirch, and M. Althoff, "Set-based prediction of pedestrians in urban environments considering formalized traffic rules," in *IEEE Int. Conf. on Intelligent Transportation Systems*, 2018.

[22] P. Zechel, R. Streiter, K. Bogenberger, and U. Göhner, "Assumptions of lateral acceleration behavior limits for prediction tasks in autonomous vehicles," in *2019 7th International Conference on Mechatronics Engineering (ICOM)*, 2019, pp. 1–6.

[23] "OpenDRIVE Open Dynamic Road Information for Vehicle Environment 1.6," Associaton for Standardization of Automation and Measuring Systems (ASAM e.V.), 2020. [Online]. Available: https://www.asam.net/standards/detail/opendrive/

[24] A. Girard, "Reachability of uncertain linear systems using zonotopes," in *Hybrid Systems: Computation and Control*, M. Morari and L. Thiele, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 291–305.

[25] C. Le Guernic and A. Girard, "Reachability analysis of hybrid systems using support functions," in *Computer Aided Verification*, A. Bouajjani and O. Maler, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 540–554.

[26] S. Söntges, M. Koschi, and M. Althoff, "Worst-case analysis of the time-to-react using reachable sets," *2018 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1891–1897, 2018.