TECHNICAL APPENDIX

A. Scene Specification File

Fig. 12 shows the scene specification file that we use in the scene generation approach. As seen, this file provides a selection schema to select the scene variables and the sampler that they want to use for scene generation. This abstracts the user from directly interacting with the complexity of the language. In addition to the samplers discussed in the paper, we also have a manual sampling approach that will allow the user to manually type in values for the scene variables.

```
Description: Carla Scene Generation
Scene Description:
    scene_length: false
    town_description: false
    roads: false
    road_segments: true
    traffic_density: false
    sensor_faults: true
    weather:
      cloudiness: true
      precipitation: true
      precipitation_deposits: false
      sun_altitude_angle: true
      wind_intensity: false
      sun_azimuth_angle: false
      wetness: false
      fog_distance: false
      fog_density: false
Samplers:
    Manual: false
    Random: false
    Halton: false
    Grid: false
    Random_Neighborhood_Search: true
    Guided_Bayesian_Optimization: false
Scene Specification:
    scene1:
        scene_length: 20
        weather:
          cloudiness: 0
          precipitation: 0
          precipitation_deposits: 0
          sun_altitude_angle: 90
          wind_intensity: 0
          sun_azimuth_angle: 0
          wetness: 0
          fog_distance: 0
          fog_density: 0
        road_segments: 2
        traffic_density: 10
```

Fig. 12: The specification file that we use with the scene generation approach.

B. ReSonAte Risk Estimation

Recall that we use the ReSonAte tool to estimate the system's dynamic risk or the ReSonAte score. For this, it estimates the hazard rate λ and the likelihood of the hazard occurrence using a BTD. Fig. 13 illustrates the dynamic likelihood of a collision and the OOD detector results for two test scenes. Here, the risk estimator and the OOD detectors

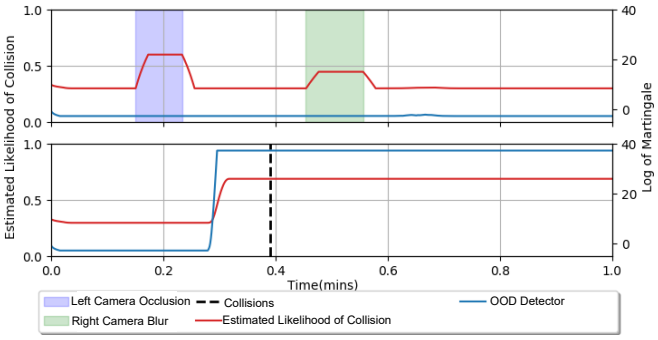

Fig. 13: The estimated collision rate for two test scenes. (Top) Scene1 - A scene with two camera faults. The average likelihood of collision across the scene was 0.4. (Bottom) Scene2 - A scene with adverse brightness. The average likelihood of collision across the scene was 0.69.

are used to estimate every simulation step. To explain, **Scene1** is a normal train scene with no adverse weather conditions. But it had two camera faults of left camera occlusion and right camera blur, which increased its likelihood of a collision. The average likelihood of a collision for the scene was 0.4. Also, the AV did not perform any infractions in this scene. **Scene2** is a scene with high brightness. As seen, the likelihood of collision is high throughout the scene because the OOD detector identified the scene to be OOD. The average likelihood of a collision for the scene was 0.69, and the system had an actual collision at 39^{th} second.

C. Notation Lookup

Notation	Description
s	scene
s_v	scene variables
v_d	variable distribution
v_c	variable constraints
$\mathbb{E}$	Environmental variables
$\mathbb{S}$	Structural variables
$\mathbb{F}$	Fault variables
$\mathcal{SC}$	Sampling Constraints
λ	dynamic hazard rate
S_{Risk}	Risk Score
RS	ReSonAte Score
IS	Infraction Score
δ	Risk threshold computed across calibration set
$\mathcal{E}$	List of previously explored scenes
k	RNS sampler's parameter to control exploration vs. exploitation
β	GBO sampler's parameter to control exploration vs. exploitation
P	Precipitation
C	Cloudiness
T	Time-of-day
RS	Road Segments
TD	Traffic Density

TABLE III: Symbols used in the paper