University of BRISTOL

Dr. Greg Chance, PhD, CEng, MInstP

Dr. Greg Chance
Trustworthy Systems Lab
University of Bristol
1 Cathedral Square
Bristol, BS1 5DD, UK
Email: greg.chance@bristol.ac.uk
URL: bristol.ac.uk

August 10, 2021

Prof. Eskandarian
Department Head
Nicholas and Rebecca Des Champs Chair
Mechanical Engineering Department, Virginia Tech
635 Prices Fork Road, 449 Goodwin Hall (MC 0238) Blacksburg, VA 24061

Review of paper T-ITS-21-05-1111.

Dear Prof Azim Eskandarian,

Thank you for taking the time to review our manuscript. We have given each comment careful consideration between all the authors and we have made changes to the document that make the document clearer and more correct. We give a full account of the changes and rebuttals below and we include a pdf diff to show the changes made.

Reviewer 1
Comment 1: "the whole manuscript seems written from the perspective of Game Engines testing, which lacks practical values in the transportation area.".

Thank you for this comment, we believe this comment requires a considered response to ensure the main message of the paper comes across to the reader. Carla is based on a game engine and Carla is currently a popular choice for simulation-based vehicle testing, so this would suggest that the underlying game engine of Carla is of practical value to the community of intelligent transportation. The paper is not focused on game engines per se, but rather the implications of using those game engines for vehicle testing.

Furthermove, we state on pg.4 the differences between the requirements of game engines for gaming, and for that of AV testing. "Considering the objectives for gaming and comparing them to these for AV development and testing, there are fundamental differences. Providing game players with a responsive real-time experience is often achieved at the cost of simulation accuracy and precision. The gamer neither needs a faithful representation of reality (i.e. gamer accepts low accuracy) nor require repeated actions to result in the same outcome (i.e. gamer accepts low precision). In contrast, high accuracy and precision are necessary for AV development, testing and verification."

Carla, and other game engine based simulators, will be an entry point for many SME's and start-up companies looking to develop products and services in this area and we believe that this paper brings pertinent information to this community, many of whom may look to ITS for guidance.

Given the confusion to reviewer 1 of our message, we believe that the title of the document should be changed to better reflect the content of the paper, which is primarily concerned with the importance of determinism for simulation based verification of autonomous vehicles. Hence we have changed the paper title to: "On the Importance of Determinism for Simulation-based Autonomous Vehicle Verification using a Game Engine".

Comment 2: "One major concern is index selection. As we all know, the scale of deviation relies on the mean values of the investigated variables. Thus, I wonder why the authors pick maximum deviation to measure the performance of the simulation results. In my opinion, the average deviation seems better in measuring the overall performance for the whole simulation process.".

The issue we found with taking the average deviation over many (100's or 1000's) of repeated runs is that a single 'failure' can be hidden in an average. For example 1 error in 1000 would be an insignificant difference to an average, but a single simulation run that exceeds permisible variance could result in a false negative result or even fail to detect the presence of a serious fault with the system under test. The verification process needs to be aware of any failure, as even a single failure may coincide with a bug or error in the system that needs to be found and corrected. Hence, it is imperative that we discover any simulation that exceeds the simulation variance. We could reframe this as thinking less about simulation performance, and more about detecting any errors in the system. Verification requires the detection of edge cases (rare cases) not the average performance of the system.

We have added a sentence to "Initial testing [26] indicated an actor path deviation of1x10-13cm for 997 out of 1000 tests, with three tests reporting a deviation of over 10cm. *Due to this nature of fault rates, using an average simulation variance would 'hide' these errors and hence, this is the reason that we use maximum variance and not an average.*".

Thank you and the reviewers for your time and consideration into this paper.

Sincerely,

Dr. Greg Chance, PhD, CEng,
MInstP