Trustworthy Autonomous Systems (TAS): The Verifiability Approach

Mohammad Reza Mousavi

Mohammad.mousavi@kcl.ac.uk

https://orcid.org/0000-0002-4869-6794

King's College London, London, UK

https://doi.org/10.26439/ciis2022.6063

Abstract

Autonomous systems are taking over the decision-making in many crucial aspects of our lives. Having the right level of trust in them will help their users benefit from such systems without harming themselves. Establishing the right level of trust involves a holistic validation and verification process, accounting for aspects such as interactions with the physical world and human users. In this talk, I present our ongoing effort in providing a holistic framework for ensuring the verifiability of autonomous systems.

Keywords: Autonomous systems, trust, verifiability, validation and verification, testing.

Sistemas Autónomos Confiables (TAS): El Enfoque de la Verificabilidad

Resumen

Los sistemas autónomos se están haciendo cargo de la toma de decisiones en muchos aspectos cruciales de nuestras vidas. Tener el nivel adecuado de confianza en ellos ayudará a sus usuarios a beneficiarse de dichos sistemas sin dañarse a sí mismos. Establecer el nivel adecuado de confianza implica un proceso holístico de validación y verificación, que tiene en cuenta aspectos como las interacciones con el mundo físico y los usuarios humanos. En esta charla, presento nuestro esfuerzo continuo para proporcionar un marco holístico para garantizar la verificabilidad de los sistemas autónomos.

Palabras clave: Sistemas autónomos, confianza, verificabilidad, validación y verificación, testing.

Cómo citar

Mousavi, M. R. (2022). Trustworthy Autonomous Systems (TAS): The Verifiability Approach. *Actas del Congreso Internacional de Ingeniería de Sistemas 2022: Entornos híbridos en la pospandemia: posibilidades para las nuevas tecnologías*, e6063. https://doi.org/10.26439/ciis2022.6063

Introduction

Autonomous systems are the result of an integration of software, hardware, and communication systems that enables decision-making with minimal intervention required from their users *(Mousavi et al. 2022)*. Examples of such systems include pacemakers and implantable defibrillators, drones and unmanned aerial vehicles (UAVs), and chatbots. Although decision making in such systems is performed autonomously, they often engage in patterns of interactions with users and hence, their usefulness crucially depends on a smooth orchestration of these interactions.

Trust and trustworthiness are a crucial aspect in the development and deployment of autonomous systems: it concerns with the users' belief that the system is going to be helpful and safe in challenging scenarios *(Araujo et al. 2019)*. Trusting a system that is not trustworthy can harm the users, since the users will adapt the systems in challenging scenarios that the system cannot cope with. Likewise, not trusting a system that is trustworthy can lead to avoiding the system in scenarios that the system can cope with and hence, not benefitting from the system. Establishing the right level of trust involves gathering and communicating sufficient evidence for system's safety and usefulness. A holistic validation and verification process are an essential ingredient for providing such evidence *(Mousavi et al. 2022; Araujo et al. 2022)*.

In this talk, I will go through our verifiability framework for autonomous systems. This involves 1) learning about system and user behaviour and capturing their appropriate models *(Damasceno et al. 2021)*; 2) adapting the models by observing and adapting to changes *(Damasceno et al. 2019; Tavassoli et al. 2022)*; 3) generating structured test suites that cover different aspects of system and user behaviour *(Araujo et al. 2020; Biewer et al. 2022)*; and 4) an analysis of the test results and explaining the patterns of interaction *(Sarda Gou et al. 2022)*.

For each of the above-mentioned four steps, we review our latest results, and point out the challenges before us in establishing a holistic verification framework for autonomous systems.

References

Araujo, H. L. S., Damasceno, C. D. N., Dimitrova, R., Kefalidou, G., Mehtarizadeh, M., Mousavi, M. R., Onime, J., Ringert, J. O., Rojas, J. M., Verdezoto, N. X., & Wali, S. (2019). Trusted Autonomous Vehicles: An Interactive Exhibit. 2019 IEEE International Conferences on Ubiquitous Computing & Communications (IUCC) and Data Science and Computational Intelligence (DSCI) and Smart Computing, Networking and Services (SmartCNS), 386–393. https://doi.org/10.1109/IUCC/DSCI/SmartCNS.2019.00091

Araujo, H., Hoenselaar, T., Mousavi, M. R., & Vinel, A. (2020). Connected Automated Driving: A Model-Based Approach to the Analysis of Basic Awareness Services. 2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications, 1–7. https://doi.org/10.1109/PIMRC48278.2020.9217142

Araujo, H., Mousavi, M. R., & Varshosaz, M. (2022). Testing, Validation, and Verification of Robotic and Autonomous Systems: A Systematic Review. ACM Trans. Softw. Eng. Methodol. https://doi.org/10.1145/3542945

Biewer, S., Dimitrova, R., Fries, M., Gazda, M., Heinze, T., Hermanns, H., & Mousavi, M. R. (2022). Conformance Relations and Hyperproperties for Doping Detection in Time and Space. Logical Methods in Computer Science, 18(1). https://doi.org/10.46298/lmcs-18(1:14)2022

Damasceno, C. D. N., Mousavi, M. R., & da Silva Simao, A. (2019). Learning to Reuse: Adaptive Model Learning for Evolving Systems. Integrated Formal Methods, 138–156. https://doi.org/10.1007/978-3-030-34968-4_8

Damasceno, C. D. N., Mousavi, M. R., & Simao, A. da S. (2021). Learning by sampling: Learning behavioral family models from software product lines. Empirical Software Engineering, 26(1), 4. https://doi.org/10.1007/s10664-020-09912-w

Gou, M. S., Lakatos, G., Holthaus, P., Wood, L., Mousavi, M. R., Robins, B., & Amirabdollahian, F. (2022). Towards understanding causality – a retrospective study of using explanations in interactions between a humanoid robot and autistic children. 2022 31st IEEE International Conference on Robot and Human Interactive

Communication (RO-MAN), 323–328. https://doi.org/10.1109/RO-MAN53752.2022.9900660

Mousavi M. R., Cavalcanti A., Fisher M., Dennis L., Hierons R., Kaddouh B., Law E.L., Richardson R., Ringert J.O., Tyukin I., & Woodcock J (2022). Trustworthy Autonomous Systems through Verifiability. IEEE Software.

Tavassoli, S., Damasceno, C. D. N., Khosravi, R., & Mousavi, M. R. (2022). Adaptive Behavioral Model Learning for Software Product Lines. Proceedings of the 26th ACM International Systems and Software Product Line Conference - Volume A, 142–153. https://doi.org/10.1145/3546932.3546991

Bio

Mohammad Reza Mousavi is a professor of Software Engineering at King's College London. He got his bachelors and master’s degree in computer engineering and Software Engineering, respectively in 1999 and 2001, from Sharif University of Technology, Iran. Subsequently, he obtained his Ph.D. in Computer Science from Eindhoven University of Technology, The Netherlands in 2005. He held positions at Reykjavik University (postdoctoral researcher), Eindhoven University of Technology (assistant and associate professor), Delft University of Technology (guest faculty member), Halmstad University (professor of Computer Systems Engineering), Chalmers / University of Gothenburg (guest professor of Software Engineering), and the University of Leicester (professor of Data-Oriented Software Engineering). Mohammad's main research area is in model-based testing, particularly applied to software product lines and cyber-physical and autonomous systems. He has been leading several research initiatives and industrial collaboration projects on healthcare and automotive systems their validation and verification.

Biografía

Doctor en Ciencias de la Computación por la Universidad de Tecnología de Eindhoven (Países Bajos). Antes de incorporarse al King’s College de Londres en 2021, ocupó cargos en la Universidad de Reikiavik, la Universidad Tecnológica de Eindhoven, la Universidad Tecnológica de Delft, la Universidad de Halmstad, la Universidad Tecnológica de Chalmers y la Universidad de Leicester. Entre sus temas de interés

están las validaciones basadas en modelos, pruebas y verificación de sistemas ciberfísicos y líneas de productos de software de prueba. Ha liderado iniciativas de investigación y proyectos de colaboración industrial en sistemas de salud y automoción, así como su validación, verificación y certificación.