Trustworthy Artificial Intelligence for Securing Transportation Systems

Bhavani Thuraisingham
Department of Computer Science, The University of Texas at Dallas
Richardson, Texas, United States
bxt043000@utdallas.edu

ABSTRACT

Artificial Intelligence (AI) techniques are being applied to numerous applications from Healthcare to Cyber Security to Finance. For example, Machine Learning (ML) algorithms are being applied to solve security problems such as malware analysis and insider threat detection. However, there are many challenges in applying ML algorithms for various applications. For example, (i) the ML algorithms may violate the privacy of individuals. This is because we can gather massive amounts of data and apply ML algorithms on the data to extract highly sensitive information. (ii) ML algorithms may show bias and be unfair to various segments of the population. (iii) ML algorithms themselves may be attacked possibly resulting in catastrophic errors including in cyber physical systems such as transportation systems. Finally, (iv) the ML algorithms must be safe and not harm society. Therefore, when ML algorithms are applied to transportation systems for handling congestion, preventing accidents, and giving advice to the drivers, we must ensure that they are secure, ensure privacy and fairness, as well as provide for the safe operation of the transportation systems. Other AY techniques such as Generative AI (GenAI) are also being applied not only to secure systems design but also determine the attacks and potential solutions.

This presentation is divided into two parts. First, we describe our research over the past decade on Trustworthy ML systems. These are systems that are secure as well as ensure privacy, fairness, and safety. We discuss our ensemble-based ML models for detecting attacks as well as our research on developing Adversarial Machine Learning techniques. We also discuss securing the Internet of Transportation systems that is based on traditional methods such as Extended Kalman Filters to detect cyberattacks. Second, Second, we discuss our work on Finally, we discuss the research we recently started as part of the USDOT National University Technology Center **TraCR** (**Tra**nsportation **C**ybersecurity and **R**esiliency) led by Clemson University. In particular, we describe (i) the application of federated machine learning techniques for detecting attacks n transportation systems; (ii) publishing synthetic transportation data sets that preserves privacy, (iii) fairness algorithms for transportation systems, and (iv) examining how GenAI systems are being integrated with transportation systems to provide security. Our focus includes the following:

- Data Privacy: We are designing a Privacy-aware Policy-based Data Management Framework for Transportation Systems. Our work involves collecting the requisite data and developing analysis tools to identify and quantify privacy risks. Existing privacy-preserving, differentially private synthetic data generation techniques, which tailor data utility for generic ML accuracy, are not well suited for specific applications. We are developing synthetic data generation tools for transportation systems applications. We will develop new ML algorithms that can leverage these datasets.

- Fairness: We have developed a novel adaptive fairness-aware online meta-learning algorithm, FairSAOML, which adapts to changing environments in both bias control and model precision. Our current work is focusing on adapting our framework to fairness in transportation systems. and control bias over time, especially ensure group fairness across different protected sub-populations; identifying interesting attributes using explainable AI techniques that might help to mitigate bias and develop equitable algorithms. We have also developed a second system, FairDolce, that recognizes objects involving fairness constraints in a changing environment. We are adapting it to transportation applications. For example, pedestrian detection (whether or not the object being seen is a pedestrian) must be fair with respect to race or gender of the individuals being detected under changing environments (e.g., rainy, cloudy sunny). Adversarial ML: Our prior work on adversarial ML models worked on traditional datasets such as network traffic data. Our current focus is on adapting our approach to AV-based sensor data. Our ML models are being applied to sensor data for object recognition and traffic management. These ML models may be attacked by the adversary. We will study various attack models and investigate ways of how interactions may occur between the model and the adversary and subsequently develop appropriate adversarial ML models that operate on the AV sensor data.

SACMAT 2024, May 15–17, 2024, San Antonio, TX, USA

ACM ISBN 979-8-4007-0491-8/24/05. https://doi.org/10.1145/3649158.3657041

- Attack Detection – Smart vehicles are often exposed to various attacks making it difficult for manufacturers to collaboratively train anomaly/attack detection models. Yet it would be ideal if all the data available across manufacturers could be used in building robust attack detection systems. To achieve this, we developed FAST-SV, which incorporates federated learning in conjunction with augmentation techniques to build a highly performant attack detection system for smart cars.

Safety: Safety has been studied for cyber physical systems and formal methods have been applied to specify safety properties and subsequently verify that the system satisfies the specifications. However, our goal is to ensure that the ML algorithms utilized by the transportation systems are safe. This would involve developing an AI Governance framework that would require transparency and explainability (among others) of the ML algorithms utilized by the transportation system.

CCS CONCEPTS

• Computing Methodologies • Artificial Intelligence

KEYWORDS

Trustworthy Artificial Intelligence, Attack Detection, Adversarial Machine Learning, Data Privacy, Fairness, Transportation Systems

ACM Reference Format

Bhavani Thuraisingham. 2024. Trustworthy Artificial Intelligence for Securing Transportation Systems. In *Proceedings of the 29th ACM Symposium on Access Control Models and Technologies (SACMAT 2024), May 15–17, 2024, San Antonio, TX, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3649158.3657041

BIOGRAPHY

Dr. Bhavani Thuraisingham is the Founders Chair Professor of Computer Science and the Founding Executive Director of the Cyber Security Research and Education Institute at The University of Texas at Dallas (UTD (2004-2021). She was a a visiting Senior Research Fellow at Kings College, University of London 2015-2022, and an Associate Director of the USDOT National UTC on Transportation Systems Security since 2023. She is an elected Fellow of multiple organizations including the ACM, IEEE, AAAS, NAI (National Academy of Inventors) and the BCS (British Computer Society). Her research interests are on (i) integrating cyber security and data science/ML/AI including as they relate to social media, Internet of Things, and more recently Transportation Systems as well as (ii) Cyber Security and AI Policy and Governance. She is also a Co-Director of the Women in Cyber Security and Women in Data Science Centers at UTD.

Dr. Thuraisingham has received several technical, education and leadership awards including the IEEE CS 1997 Edward J. McCluskey Technical Achievement Award, the 2023 IEEE CS Taylor L. Booth Education Award, ACM SIGSAC 2010 Outstanding Contributions Award, the IEEE ComSoc Communications and Information Security 2019 Technical Recognition Award, the IEEE CS TC (Technical Committee) on Services Computing 2017 Research Innovation Award, the ACM CODASPY 2017 Lasting Research Award, the IEEE ISI 2010 Research Leadership Award, the ACM SACMAT 10 Year Test of Time Awards for 2018 and 2019 (for papers published in 2008 and 2009) and the IEEE CS Cloud TC on Women in Cloud Computing Award. Her 43+ year career includes industry (Honeywell), Federal Research Laboratory (MITRE), US government (NSF) and Academia. She has delivered over 200 keynote/featured addresses, over 100 panel presentations including at the United Nations, White House Office of Science and Technology Policy, Fortune Media, Dell Technologies World, Lloyds of London Insurance, and Professors Without Borders, written 16 books, published over 130 journal articles (including in several IEEE and ACM Transactions) and over 300 conference papers (including in top tier venues such as ACM CCS, ACM KDD, IEEE ICDM, IEEE ICDE, and AAAI), has 7 US patents and participated in several podcasts. She has also written opinion columns on security for venues such as the New York Times, WomensDay.com, Inc. Magazine, the Legal 500 and the Connected World. Out of the 23 PhD students she has graduated at UTD, 11 are female and others include those from the African American, Latino American, and the LGBTQ communities. She also works with neurodiverse students in cyber security. Over the years she has educated a global community and is affiliated with the University of Dschang in Cameroon, Africa since 2021 and educates the general public in practicing cyber hygiene including talks at public libraries in DFW.

Dr. Thuraisingham received her PhD from the University of Wales, Swansea, UK, and the prestigious earned higher doctorate (D. Eng) from the University of Bristol, UK. She has a Certificate in Public Policy Analysis from the London School of Economics and Political Science. She has been featured in the book by the ACM titled: "Rendering History: The Women of ACM-W" as one of the 30+ "*Women that Changed the Face of Worldwide Computing Forever.*"

ACKNOWLEDGMENTS

The research was supported in part by the National Center for Transportation Cybersecurity and Resiliency (TraCR) (a U.S. Department of Transportation National University Transportation Center) headquartered at Clemson University, Clemson, South Carolina, USA. Any opinions, findings, conclusions, and recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of TraCR, and the U.S. Government assumes no liability for the contents or use thereof.