Research Statement by Hussein Sibai

My research is on trustworthy autonomy, spanning over the areas of cyber-physical systems, formal methods, control theory, and machine learning. I develop theory and implement tools to efficiently design, evaluate, and deploy autonomous systems.

Motivation: Autonomy research has made significant progress in solving the most challenging perception, planning, and control problems. Such progress is advancing exciting applications like autonomous driving and robotic surgery. Unfortunately, the resulting systems are also known to be expensive and to have serious faults that led to fatalities. Rigorous and efficient methods for their design, evaluation, and deployment are thus in urgent need. Carefully applied formal approaches that exploit the structure inherent in these systems can transform practice and provide insights for effective fulfillment of that need. I create such approaches while drawing ideas from control theory, formal methods, testing, and machine learning.

Overview: My Ph.D. research develops techniques that reduce the cost of state estimation, formal verification, and testing of autonomous systems. I use the theory of symmetry of dynamical systems to optimize existing formal verification and testing techniques for cyber-physical systems (CPSs) [1, 2, 3, 4, 5, 6]. In the second line of work, I use the theory of topological entropy of dynamical systems to minimize the bit-rates needed for state estimation of CPSs over limited-bandwidth channels [7, 8, 9, 10]. My results were published in leading control and formal verification venues such as HSCC, ATVA, TACAS, and CAV, won the ACM graduate-category gold medal in the student research competition in CPS-IoT Week'21, were best student paper award finalists in HSCC'17 and ATVA'19, won best poster award in HSCC'18, and presented in a spotlight talk in the Negative Results Workshop in CVPR'17. My research was recognized by the Rambus, Ernest A. Reid, and Mavis Future Faculty fellowships from the University of Illinois Urbana-Champaign. I also contributed to building several autonomous driving capabilities for a real car [11], designing a simulation and verification framework for unmanned air-traffic management (UTM) protocols [12], optimizing bandwidth allocation in Tor [13, 14], and analyzing the differential privacy of sequential algorithms [15]. I collaborated with leading research groups in diverse fields, including computer vision, privacy, security, robotics, mechanical engineering, and software engineering. I collaborated with, and my research was partially funded by, The Boeing Company for the past four years. My research was also partially funded by the AFOSR and the NSA. Finally, I interned at the National Institute of Aerospace.

Three highlights from my recent research are:

1. I designed a symmetry-based counterexample-guided abstraction-refinement (CEGAR) algorithm for CPSs that enabled an average of 14× speedup in verifying the safety of autonomous vehicles (AVs) with neural network-based (NN) controllers navigating complex scenarios [1, 2, 3, 4] (Section 1.1).

2. I derived the first upper bounds on the minimal bit-rates needed for the state estimation of CPSs over limited-bandwidth channels [7, 8, 9, 10] (Section 1.2).

3. I empirically demonstrated that object detectors are robust against physical adversarial examples generated by existing methods, then demonstrated a successful attack [16, 17, 18] (Section 1.3).

On the practical side, I spent significant effort developing the symmetry-utilizing scenario verification tool SceneChecker [2]. I deployed a reachability analysis-based decision-making algorithm on an actual Polaris GEM Research and Development vehicle that ensures the safety of nearby pedestrians [11]. Finally, I developed software to evaluate UTM and Airborne Collision Avoidance System (ACAS) protocols in simulation using Robot Operating System (ROS) and Gazebo [12, 5].

1 Ph.D. research on verification, state estimation, and robust perception

1.1 Accelerating verification of CPSs using symmetry

Formal verification of CPSs is often based on *reachability analysis*, where the *reachset*, the set of states that the system can reach, has to be computed. In most cases, exact reachset computation is undecidable, and its approximation is computationally expensive. A line of research spanning over three decades has focused on building data structures, algorithms, and software tools to optimize approximate reachset computations in terms of accuracy and computational cost. Scalability remains a major challenge.

I wondered how further knowledge about the structure of a CPS could be utilized for more efficient reachability analysis. Common structural property is symmetry. A symmetry of a dynamical system is a map acting on its state space that transforms any of its trajectories to another one of its trajectories. For example, the translation of the initial position of a car on a straight road would not change the shape of the trajectory but would merely translate it. A symmetry then transforms a computed reachset to another reachset. Such a transformation is cheaper to execute than existing reachset-computation methods.

Using symmetry for caching reachsets: Simulation-based verification algorithms implemented in tools such as DryVR and C2E2 compute accurate reachsets of a CPS by recursively partitioning the initial set of states and computing the reachset of each part through simulations and sensitivity analysis. I designed a tree-shaped cache to store and retrieve the multi-fidelity reachsets being computed and used symmetries to increase the cache-hit rate. I implemented the cache in DryVR. It achieves up to a two orders-of-magnitude speedup in verification time for several scenarios. The scenarios include a collision-check of two AVs sharing the same lane [4]. I extended this approach to hybrid automata, a modeling framework for CPSs with continuous and discrete dynamics. I created a cache that shares reachset computations of the system at different discrete states, or *modes*, with symmetric continuous dynamics. My colleague Navid Mokhlesi and I implemented this idea for faster safety verification of multi-agent systems. Our experiments show up to 64% speedup in verifying the safety of a fixed-wing plane swarm using translation, rotation, and permutation symmetries [3].

Symmetry abstractions: I designed an algorithm that reduces the size of the initial set of states of a CPS, starting from which its safety needs to be verified. The algorithm combines the initial set's symmetric states before computing the reachset. This algorithm saves cache-access time and, in some instances, achieves a dimensional reduction of the initial set. When verifying the same scenarios, it achieves up to 10× additional speedup over the caching-based algorithm discussed above [4]. Similarly, I designed an abstraction algorithm for hybrid automata that preemptively groups their modes that possess symmetric continuous dynamics together, creating simpler ones that are easier to verify [1]. I developed a corresponding CEGAR algorithm that automatically refines an unsafe abstract automaton. If verifying an abstract automaton resulted in a witness counterexample execution as proof for non-safety, the algorithm checks if the counterexample is valid, i.e., corresponds to an actual counterexample execution of the concrete automaton. The algorithm selectively splits some of the abstract modes if the counterexample is not valid, eliminating it from the executions of the abstract automaton. The algorithm efficiently ensures that using the abstraction does not affect the accuracy of the verification result. Finally, my colleague Yangge Li and I implemented a simplified version of the algorithm in a novel open-source Python tool called SceneChecker. Our experiments evaluating SceneChecker show a 14× average speedup in verifying 2D and 3D scenarios where cars and drones with NN-based controllers follow sequences of hundreds of waypoints in complex environments with hundreds of obstacles [2].

My methods can be integrated to speed up any existing reachability analysis-based verification tools for the wide range of CPSs that possess symmetries, most importantly AVs. They can accelerate the generation of the operating volumes for aircraft during the execution of the UTM protocol [12].

1.2 State estimation of CPSs under bandwidth constraints

Modern dynamical systems are estimated and controlled over digital channels with limited bandwidth. Optimizing the design of these channels requires the knowledge of the minimal bit-rates needed to estimate and control given error specifications. Researchers like Colonius, Kawan, Zamani, Liberzon, and Mitra, have presented different notions of topological entropy and shown that they lower-bound the minimal bit-rates needed for different tasks, including stabilization, invariance control, and state estimation. Topological entropy for a dynamical system is the rate of the exponential increase of the number of its distinguishable trajectories. I proposed new entropy notions that lower-bound the bit-rates needed for the state estimation of closed nonlinear switched systems and nonlinear continuous-time systems with inputs, common CPSs. I derived corresponding entropy upper bounds and designed bandwidth-efficient state estimation algorithms [8, 10]. I showed that my two entropy notions are closely related while previously existing notions have infinite values for simple systems, making them unsuitable for the types of systems I considered [7, 9]. These results contribute to constructively answering the minimal-bit-rate questions for state estimation of widely deployed CPSs, laying the foundations for designs with lower bandwidth costs and power consumption. Such costs significantly limit the millage of AVs, for example, and hinder their deployment.

1.3 Robustness of object detectors against physical adversarial examples

Adversarial examples are inputs to machine-learning (ML) models that have been perturbed with humanly-imperceptible noise, causing dramatic unwanted changes of their output. In 2017, there was an abundance of methods to attack image classifiers with digitally-perturbed images, but there were only a few to do that by physically perturbing the pictured objects instead. I was curious if these methods can trick NN-based object detectors, such as YOLO and Faster-RCNN, instead of classifiers, to miss-recognize the resulting physically-perturbed objects in most of the frames of videos taken of them from different angles, distances, and environmental conditions, thus becoming a fatal danger. The variety of capture conditions alters the physically-added adversarial noise and may lessen its effectiveness against the detectors. Additionally, an object detector does not directly classify the objects. It first has to locate them in the image and crop their bounding box. The imperfection of the localization and cropping process makes it harder to attack detectors compared to classifiers. My collaborators, Jiajun Lu, Evan Fabry, and David Forsyth, and I empirically checked if existing methods at the time would cause a miss-recognition of a perturbed stop sign in a video taken from an approaching vehicle. Our experiments showed that the perturbations were not robust to the adjustments made by the printer, camera, object detector, and capture distances and angles [16, 18]. We improved existing methods and constructed robust perturbations, but at the expense of being perceptible [17]. Eykholt et al. later introduced similar successful perturbations on graffiti-like posters glued to stop signs demonstrating realizable attacks. Since then, more advanced attacks, defenses, and verification techniques have been developed for ML models and the systems incorporating them, influencing their design, evaluation, and deployment. Many fascinating questions remain open.

2 Future research plan

The recent incorporation of ML components for perception and control in CPSs has brought game-changing opportunities and challenges in their design, evaluation, and deployment. The increasing development of connected CPSs raises concerns about their security, privacy, reliability, and power consumption. My research plan aims to tackle these fundamental problems. Symmetry, topological entropy, and bounded sensitivity to perturbations are structural properties that describe overlapping aspects of similarity and complexity of the trajectories of a CPS. Other similar properties include stability, decomposability into interconnected smaller subsystems, monotonicity, differential flatness, and having a causal structure. All such properties have the

immense potential to help address the problems above effectively. Other types of knowledge about a CPS with similar prospects are data, probability distributions, mathematical models, and simulators. I am eager to collaborate with passionate researchers from interdisciplinary fields and sectors to design algorithms that harness these potentials and deploy them to achieve safe and efficient CPSs.

Efficient design, evaluation, and deployment of CPSs with ML components: The emergence of ML models that are capable of analyzing inputs with many dimensions, e.g., images and point clouds, or approximating complex controllers, e.g., model-predictive ones, is from the key enablers of ambitious CPSs such as AVs. Promising formal certification methods and tools for such models and the systems incorporating them have been developed in the past five years. However, these methods are, so far, only capable of checking the satisfiability of local properties of a few small regions in the input and state spaces of simple academic examples. Testing using high-fidelity simulators is another prominent approach for a faster and risk-free evaluation of CPSs. However, such simulations are still slow and can become prohibitive, mainly for regression testing. Similar scalability challenges are faced in the design process at the model selection, training, and system integration phases. At deployment, especially at the edge, power consumption and delay problems arise because of the demanding computations of these models. As the traditional CPSs, ML-enabled CPSs possess structural properties worth utilizing. For example, growing research on graph NNs, equivariant NNs, causal ML, and symbolic artificial intelligence (AI) incorporate domain knowledge in ML models. I aim to effectively utilize or enforce the structure in these models corresponding to such knowledge to design, evaluate, and deploy safe and efficient ML-enabled CPSs. As a start in this thrust, I developed an efficient testing algorithm that utilizes user-provided symmetries of the continuous dynamics of multi-agent CPSs to accelerate the execution of their corresponding high-fidelity test suites [5]. I applied it to test an ACAS protocol implemented in a Julia-based executable and integrated with Hector quadrotors in Gazebo.

Optimizing bandwidth for private and robust estimation and control: In contrast to the idealistic assumptions in the research on control under bandwidth constraints, channels might be noisy, a state might be only partially observable, sensors might be defective or adversarial, or part of the state might be private or irrelevant. I plan to find the minimal bit-rates necessary to estimate or control dynamical systems in such settings. The bit-rates would depend on the complexity of the system dynamics and on parameters that determine the state's observability, e.g., the state-output map, number of adversaries, and the added noise. For example, in the case of adversaries or sensor failures, necessary bit-rates increase for error correction, while in private settings, they decrease for information hiding. I plan to derive theoretical bounds on such bit-rates and design bandwidth-efficient estimation and control algorithms for such settings. This project bridges the fields of limited-bandwidth control, distributed systems, coding theory, differential privacy, sensor placement, and robust control. It aims to optimize the bandwidths used by Internet-of-Things (IoT) devices, networked CPSs, and vehicle-to-vehicle (V2V) communications while maintaining privacy and robustness.

Building trustworthy autonomous systems: Alongside my strong interest in building theoretical frameworks for design and verification, I am equally interested in ensuring their applicability in real life, addressing the Sym2Real question. I implemented a safe decision-making algorithm for the GEM AV platform based on reachability analysis to ensure the surrounding pedestrians' safety [11]. I assisted in developing its perception and control modules. I experienced the gap between theory and practice firsthand, e.g., in error models, synchronization issues, and power constraints. I plan to build or contribute to an assured autonomy laboratory with various systems, including IoT devices, drones, race cars, and potentially an AV platform. Such a setting would offer immense information about overlooked but critical assumptions made in theoretical developments. It will also be beneficial to provide students with hands-on experience.

References

[1] Hussein Sibai and Sayan Mitra. Symmetry counter-example abstraction-refinement (cegar) algorithm for verifying hybrid automata, (under preparation).

[2] Hussein Sibai, Yangge Li, and Sayan Mitra. Scenechecker: Boosting scenario verification using symmetry abstractions, CAV'21.

[3] Hussein Sibai, Navid Mokhlesi, Chuchu Fan, and Sayan Mitra. Multi-agent safety verification using symmetry transformations. TACAS'20.

[4] Hussein Sibai, Navid Mokhlesi, and Sayan Mitra. Using symmetry transformations in equivariant dynamical systems for their safety verification. ATVA'19.

[5] Hussein Sibai, Yangge Li, Reyhan Jabbarvand Behrouz, and Sayan Mitra. Efficient testing of autonomous systems by exploiting symmetries: A case study on airborne collision avoidance systems (under preparation).

[6] Sayan Mitra and Hussein Sibai. Symmetry for boosting algorithmic proofs of cyber-physical systems (to appear). IEEE Computer Magazine (Algorithms column), 2022.

[7] Hussein Sibai and Sayan Mitra. State estimation of open dynamical systems with slow inputs: entropy, bit rates, and relation with switched systems, (conditionally accepted in IEEE TAC).

[8] Hussein Sibai and Sayan Mitra. State estimation of dynamical systems with unknown inputs: entropy and bit rates. HSCC '18.

[9] Hussein Sibai and Sayan Mitra. Poster: Recent results in state estimation of dynamical systems with inputs under bandwidth constraints. HSCC '18.

[10] Hussein Sibai and Sayan Mitra. Optimal data rate for state estimation of switched nonlinear systems. HSCC '17.

[11] Peter Du, Zhe Huang, Tianqi Liu, Ke Xu, Qichao Gao, Hussein Sibai, Katherine Driggs-Campbell, and Sayan Mitra. Online monitoring for safe pedestrian-vehicle interactions. ITSC'20.

[12] Chiao Hsieh, Hussein Sibai, Hebron Taylor, Yifeng Ni, and Sayan Mitra. Skytrakx: A toolkit for simulation and verification of unmanned air-traffic management systems. ITSC'21.

[13] Hussein Darir, Hussein Sibai, Nikita Borisov, Geir Dullerud, and Sayan Mitra. Tightrope: Towards optimal load-balancing of paths in anonymous networks. WPES'18.

[14] Hussein Darir, Hussein Sibai, Chin-Yu Cheng, Nikita Borisov, Geir Dullerud, and Sayan Mitra. Mleflow: Learning from history to improve load balancing in tor. PETS'22.

[15] Yu Wang, Hussein Sibai, Sayan Mitra, and Geir E. Dullerud. Differential privacy for sequential algorithms, arXiv'20.

[16] Jiajun Lu*, Hussein Sibai*, Evan Fabry, and David Forsyth. NO need to worry about adversarial examples in object detection in autonomous vehicles. Workshop on Negative Results, CVPR'17.

[17] Jiajun Lu*, Hussein Sibai*, and Evan Fabry. Adversarial Examples that Fool Detectors. arXiv'17.

[18] Jiajun Lu, Hussein Sibai, Evan Fabry, and David Forsyth. Standard detectors aren't (currently) fooled by physical adversarial stop signs. arXiv'17.