Safe Autonomy: Verification and Synthesis Algorithms

Research Statement by Chuchu Fan

Vision: Recent decades have witnessed a phenomenal level of investment in safe autonomous systems such as self-driving cars, drones, and medical devices. Designers of these complex systems cannot foresee all corner-cases that may arise in the field, and a single design defect can wreak havoc across thousands of deployed instances. Current approaches to ensure safety mainly rely on large-scale simulations and field tests, in an iterative bug-fixing process. There are two problems with that brute-force methodology: cost and coverage. Simulating every scenario in a set of combinatorial choices is expensive; field tests are even more so. For example, it is estimated that billions of miles of test-driving is necessary in order to reduce the catastrophic failure rates to less than one per hour [14]. This figure is prohibitive even for large corporations. Second, it is inevitable that certain scenarios will remain untested, particularly those outside the so-called *operational design domains*. Such uncertainties grow as systems get equipped with machine learning (ML) algorithms, and the test coverage problem gets exacerbated.

I believe that rigorous approaches based on formal methods and control theory can fundamentally transform the conventional trial-and-error paradigm and improve safety in autonomous systems. Rigorous approaches can, in principle, generate provably correct decision systems, provide safety guarantees, and perform root-cause analyses. The common claim that formal techniques do not scale beyond academic problems has been falsified by recent results. **My research goal is to fulfill that transformation for designing and analyzing real-world autonomous systems, by developing computationally efficient formal techniques that can provide useful coverage at an acceptable cost.**

Background and contributions: Within the broad area of safe autonomy and cyber-physical systems (CPS), my Ph.D. research primarily spans two themes: *automatic formal verification* and *control synthesis*. *Verification* aims to provide a system with *safety proofs* or *counter-examples* that demonstrate potential defects. *Control synthesis* produces *correct-by-construction controllers* for a system to meet requirements or a proof that such a controller does not exist.

Verification and synthesis problems are challenging and known to be theoretically undecidable for typical real-world models [10]. Approximate solutions are also difficult to compute due to intractable peculiarities of CPS, including nonlinear dynamics, networked structures, and the non-deterministic and hybrid nature of models. My work builds on fundamental concepts from *formal methods* and *control theory*, aimed to address the above challenges. The **key contributions** of my thesis research are as follows.

1. Advanced the state-of-the-art on verification of CPS by developing *the first bounded safety verification algorithm* for *nonlinear hybrid systems*. This data-driven algorithm is *locally optimal* in data usage [4], and can be used for *compositional verification* of *networked autonomous systems* and systems with communication delays. Therefore, it allows us to verify large models that were previously intractable [3, 11] (Sec. 1.1).
2. Developed *the first framework* for verifying real-world CPS for which certain parts may not have precise mathematical models [8]. The key idea is to see such systems as a *white-box automaton* with embedded *black-box simulators*. With this new view, our verification approach can bring together worst-case formal reasoning on the automaton with *probabilistic reasoning* on the black-boxes (Sec. 1.2).
3. Developed an algorithm that significantly improves the practical efficiency of control synthesis for *large linear systems with disturbances* [5]. The algorithm achieves scalability by reducing the synthesis problem to *satisfiability over quantifier-free linear arithmetic* and leveraging modern SMT solvers (Sec. 1.3).

Broadly, all the above results have theoretical guarantees such as *soundness, precision, and completeness*. On the practical side, I have devoted much effort to develop software tools for these techniques: *C2E2* [9] (for verification of hybrid systems), *DryVR* [15] (for verification of systems with black-box components), and *RealSyn* [5] (for synthesis). These tools have gained momentum. For instance, C2E2 was the first tool to successfully verify the Toyota powertrain control system [2]; currently, it is also the only tool that can handle highly nonlinear models such as mixed-signal circuits [6]. DryVR has been successfully used in verifying autonomous driving maneuvers [8] and determining automotive safety integrity levels (ASIL) based on risk analysis of an advanced driver-assistance system (ADAS) feature [7]. Other researchers are also using DryVR, for example, to verify spacecraft rendezvous [1] and reinforcement learning-based planners.

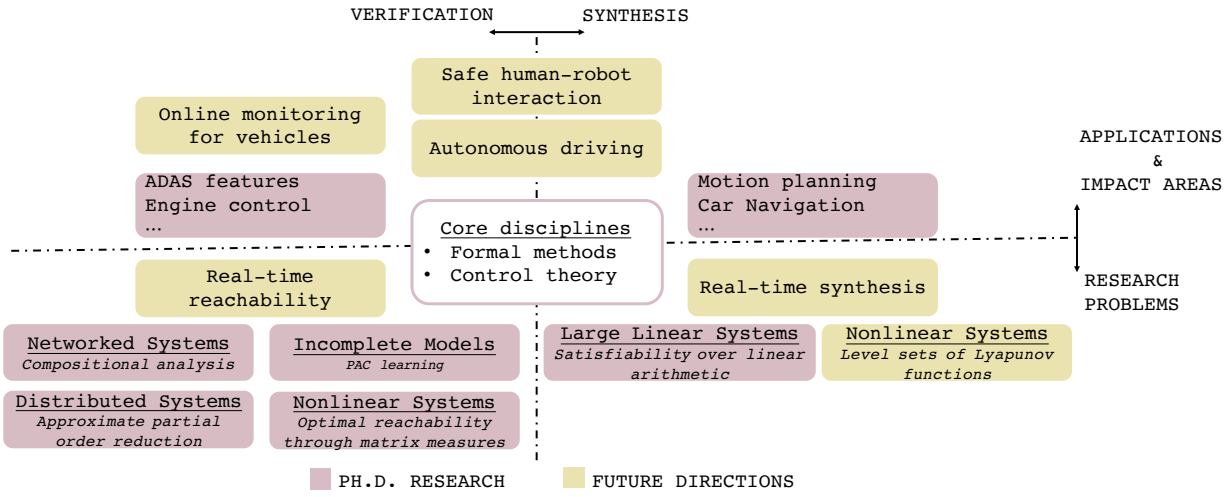

Figure 1: Overview of my Ph.D. research (red) and future plans (yellow) around the theme of safe autonomy and CPS. Formal methods and control theory in the middle are the core disciplines. Topics on the left and right are related to verification and synthesis, respectively. Basic research problems are at the bottom and applications are at the top.

1 Ph.D. Research on Formal Methods for Autonomous Systems

1.1 Sensitivity Analysis for Data-driven Verification

Unlike testing-based methods that cannot prove absence of bugs, verification provides a mathematical proof of the safety of all possible (often infinitely many) behaviors of a system. However, purely model-based verification techniques cannot directly handle practical autonomous systems that have nonlinear and hybrid models, or systems for which we lack complete or precise models. The *data-driven verification* approach that I work on combines executions of the model with model-based *reachability analysis*, by performing efficient *sensitivity analysis* for the complex or unknown components of the system. Such a sensitivity analysis gives probabilistic or worst-case bounds on how much the states or outputs of a module will change, with small changes in the input. Sensitivity analysis is the key principle underlying the rigorous soundness (i.e., the result returned by the algorithm is correct) and completeness (i.e., the algorithm always terminates and returns the correct result) guarantees of the data-driven verification approach. My work has advanced the verification of (1) autonomous systems with nonlinear dynamics [4], (2) networked autonomous systems with communication delays [11], and (3) distributed autonomous systems whose components take actions concurrently [3].

Locally optimal guaranteed reachability for nonlinear dynamics: Obtaining an exact solution for autonomous systems with nonlinear dynamics is often impossible. Nevertheless, approximate solutions for such dynamics may be too conservative or computationally expensive to be useful. Therefore, it is crucial for verification methods to perform tight yet efficient over-approximations. **I proposed the first algorithm with locally optimal guarantees for computing a tight reach sets over-approximations for nonlinear models**. I showed that the sensitivity of a system can be bounded by *matrix measures* of nonlinear dynamics, which can be computed via semi-definite programming. This allows us to develop efficient algorithms for automatic construction of the tightest reach sets. I implemented those techniques in the award-winning tool C2E2 [9], which, in turn, has allowed C2E2 to be well-recognized as a leading verification tool used by numerous research groups around the world. The success of C2E2's verification power has also attracted supports from national agencies including NSF and the Air Force for the commercialization of C2E2.

Compositional analysis of networked autonomous systems: For large-scale autonomous systems in which multiple components communicate with each other (e.g., power grids, swarm robots, and embedded medical devices), the number of state variables is the summation of those from all components, which quickly goes beyond the capability of even state-of-the-art techniques. In [11], I introduced a technique for analyzing the sensitivity of each component with respect to the changes in both initial conditions and input signals. **Based on such sensitivity, we can construct a reduced model whose dimensionality is only equal to the number of components in the system (independent of the individual**

components' dimensionality). We proved that the executions of the reduced model upper-bound the sensitivity of the original system, even if the communication between components has time delays. Building on this technique, we were able to verify a suite of challenging pacemaker-heart models that have as many as 20 continuous variables and 29^5 discrete modes [12, 13].

Approximate partial order reduction for distributed autonomous systems: In distributed autonomous systems, components take actions concurrently. Therefore, in considering all possible behaviors of a system, there is a combinatorial explosion in the total number of action sequences due to the interleavings of each individual system's concurrent actions. Existing *partial order reduction* (POR) methods are limited when it comes to computations with numerical data since reduction is allowed only when actions can commute exactly. I developed an approximate POR method that allows actions to be *nearly commutative*. **The resulting algorithm reduces the number of action sequences that must be explored in safety analysis by a factor of $O(t!)$ with t being the time steps [3].** These preliminary results have shown great promise for exponentially expediting the safety analysis of distributed systems.

1.2 Verification of Black-box Components with Probabilistic Guarantees

Many autonomous systems are a heterogeneous mix of simulation code, differential equations, block diagrams, and handcrafted look-up tables, with the increasing presence of machine learning modules. It is sometimes impossible even to model a system completely and precisely in the first place. To overcome this issue, **I presented a novel verification framework DryVR [8, 15] that treats the system as a combination of a "white-box" control graph and "black-box" simulators.** Using the *probably approximately correct* (PAC) learning principle, we showed that sensitivity analysis could be formulated as the well-known problem of learning a linear separator, and could thus be solved with probabilistic correctness guarantees. More excitingly, to achieve an error of ε with probability $1-\delta$, the number of samples the algorithm needs from the black-box simulator is only $\frac{1}{\varepsilon}\log\frac{1}{\delta}$. That approach can achieve the same level of probabilistic guarantees as testing, in significantly less time. DryVR has been used to verify a wide range of applications, including autonomous driving maneuvers and automatic transmission control [8]. We also used it to conduct risk analysis of an ADAS system to determine its ASIL [7]. Moreover, DryVR can be incorporated to enhance other research works, for example, to verify spacecraft rendezvous [1] and plan safe actions in reinforcement learning.

1.3 Control Synthesis of Large-dimensional Systems

Current control synthesis approaches suffer from poor scalability: They normally end up solving a nonlinear or mixed-integer optimization problem, or facing the curse of dimensionality. I proposed a novel approach that finds a *state feedback controller* for piecewise affine systems under disturbances with reach-avoid specifications [5]. The instrumental idea is to use a combination of an *open-loop controller* and a *tracking controller* to **reformulate the overall synthesis problem as a satisfiability problem over quantifier-free linear real arithmetic**, which can be efficiently solved by off-the-shelf SMT solvers. **The number of constraints for the satisfiability problem is only linear to the total number of surfaces in the obstacles as polytopes.** Moreover, the proposed approach is proved to be sound and complete—a theoretical guarantee that is beyond most conventional synthesis methods. The resulting tool, RealSyn, shows very encouraging results: it finds controllers within seconds for systems with up to 20 state variables, and within minutes for systems with 84 variables. I am currently working on extending this idea for control synthesis of nonlinear systems.

2 Future Research

Moving forward, I will continue working along the previously mentioned themes and also branch out to explore problems related to large-scale, multi-agent autonomous systems.

Human-robot interaction in semi-autonomous systems: Autonomous systems often interact and cooperate with humans or other agents. Consider an autonomous vehicle crossing a road with hundreds of pedestrians in New York. If it assumes that the pedestrians can do anything possible, then perhaps its only safe option is to stay still. On the other hand, rule-based planning may lead to mishaps since pedestrians might cross the road when the traffic light is red. Autonomous systems have to interact with other agents to conservatively but precisely detect their intentions. That involves **modeling**

the behaviors of humans, **modeling the communication between humans and autonomous systems**, and **developing algorithms for autonomous systems such that they can successfully achieve their goals while preserving safety**. Some of my initial ideas for approaching the problem are centered at modeling this problem as a two-player game with imperfect information. In such a game, the choice of actions for each player depends upon some information which is only partially accessible to the opponent player. We proposed an algorithm that dynamically synthesizes a winning strategy for Player A and adapts it as the game unfolds and reveals more information about the opponent player. In my future research, **I want to study how to capture the probabilistic nature of human behavior, how to model the decision-making process of humans as an optimization, how to design interactive features in an autonomous system, and how to develop formal techniques on these models to guarantee the safety and success of an autonomous system**. I also plan to implement these approaches on robots and autonomous cars. I am interested in how to tackle possible problems that can happen in realistic systems, such as noisy sensor data and delayed communications.

Online monitoring through real-time reachability: Formal methods are usually performed offline and ask for specific models and requirements. However, in reality, there are always unforeseen circumstances, so the model and requirements for autonomous systems are always evolving. **I hope to bring real-time reachability analysis together with robust control design to improve autonomous systems' ability to tolerate failures**. Take self-driving cars as an example. Each vehicle needs to perform a real-time safety analysis based on models that capture its own behavior, estimations of other agents on the road, and current traffic scenario and environment data sensed by the vehicle. Real-time reachability would determine whether the system is safe for the next T units of time, within a time that is significantly less than T. Such analysis can monitor the safety status, raise an alarm before any dangerous situation might happen, and switch to a control policy that can robustly handle potential failures based on current situations. Several techniques that I am working on hold promise for offering real-time reachability analysis. I plan to parallelize the data-driven verification approach and exploit approximate POR to expedite the computation of reach sets. However, several other problems may arise even if we can perform real-time reachability. For example, a coarse over-approximation of reach sets can trigger more false alarms, but finer estimation always takes more time. I am interested in these theoretical problems, e.g., how to optimize the trade-off between the quality and efficiency of monitoring.

Multi-robot systems with real-time synthesis: Next generation autonomous systems will require the cooperation among increasing numbers of agents. This vision calls for a suite of real-time planning algorithms for multi-robot systems, whose scale is beyond the scope of most current synthesis technologies. My work on synthesis [5] has shown promise for performing real-time synthesis for large-scale linear systems; our prototype tool can synthesize controllers within 0.1 second for a queue of 10 vehicles to maintain a platoon for the next 10 units of time. There are a lot of interesting questions here; for example, **how can we handle more complex models and missions, and how can we deal with attacks and failures during the synthesis process**? I plan to bring in ideas from distributed systems and security, along with transportation engineering, to formulate these problems in real-time synthesis and study both their theoretical and experimental aspects. In a broader context, this work fits into the concept of smart cities as symbiotic autonomous systems, in which agents coordinate to pursue their goals and ensure macro-level performance.

References

[1] N. Chan and S. Mitra. Verified hybrid lq control for autonomous spacecraft rendezvous. In *CDC*, 2017.
[2] P. S. Duggirala, C. Fan, S. Mitra, and M. Viswanathan. Meeting a powertrain verification challenge. In *CAV*, 2015.
[3] C. Fan, Z. Huang, and S. Mitra. Approximate partial order reduction. In *FM*. 2018.
[4] C. Fan, J. Kapinski, and X. Jin. Locally optimal reach set over-approximation for nonlinear systems. In *EMSOFT*, 2016.
[5] C. Fan, U. Mathur, S. Mitra, and M. Viswanathan. Controller synthesis made real: Reach-avoid specifications and linear dynamics. In *CAV*, 2018.
[6] C. Fan, Y. Meng, U. Maier, E. Bartocci, S. Mitra, and U. Schmid. Verifying nonlinear analog and mixed-signal circuits with inputs. In *ADHS*, 2018.
[7] C. Fan, B. Qi, and S. Mitra. Data-driven formal reasoning and their applications in safety analysis of vehicle autonomy features. *IEEE Design & Test*, 2018.
[8] C. Fan, B. Qi, S. Mitra, and M. Viswanathan. Dryvr: Data-driven verification and compositional reasoning for automotive systems. In *CAV*, 2017.
[9] C. Fan, B. Qi, S. Mitra, M. Viswanathan, and P. S. Duggirala. Automatic reachability analysis for nonlinear hybrid models with c2e2. In *CAV*, 2016.
[10] T. A. Henzinger, P. W. Kopke, A. Puri, and P. Varaiya. What's decidable about hybrid automata? *Journal of Computer and System Sciences*, 1998.
[11] Z. Huang, C. Fan, A. Mereacre, S. Mitra, and M. Kwiatkowska. Invariant verification of nonlinear hybrid automata networks of cardiac cells. In *CAV*, 2014.
[12] Z. Huang, C. Fan, A. Mereacre, S. Mitra, and M. Kwiatkowska. Simulation-based verification of cardiac pacemakers with guaranteed coverage. *IEEE Design & Test*, 32(5):27–34, 10 2015.
[13] Z. Huang, C. Fan, and S. Mitra. Bounded invariant verification for time-delayed nonlinear networked dynamical systems. *NAHS*, 2017.
[14] P. Koopman and M. Wagner. Autonomous vehicle safety: An interdisciplinary challenge. *IEEE Intelligent Transportation Systems Magazine*, 2017.
[15] B. Qi, C. Fan, M. Jiang, and S. Mitra. Dryvr 2.0: A tool for verification and controller synthesis of black-box cyber-physical systems. In *HSCC*, 2018.