Stochastic Reachability for Safety Verification of Cyber-Physical Systems

Kendra Lesser

Index Terms

Stochastic reachability, formal verification, computational methods, imperfect information

The scale and complexity of cyber-physical systems (CPS), as well as their prevalence in safety critical and expensive applications, dictates a need for formal safety verification of such systems. In order to advance as a mature and reliable technology in areas such as smart grids, transportation, and medical devices, a theoretical framework for verification of CPS is essential. CPS must integrate nontrivial physical dynamics with hierarchical mode logic, often in the presence of uncertain or stochastic behaviors. Standard verification techniques such as model checking and deductive methods have proven useful for moderate dimensional systems, with the added advantage of decidability, but do not handle stochastic systems well, either because of the unbounded nature of some probabilistic models, or because of overly conservative results when considering bounded stochastic systems in the worst case. The concept of stochastic reachability, however, provides a flexible and meaningful framework for establishing both theoretical and computational assurances of safety in CPS.

In order to develop a theoretical framework for addressing safety verification concerns using stochastic reachability, several research challenges must be addressed. *These research challenges include: a) safety assurances in the presence of human decision-makers; b) developing efficient computational methods for high-dimensional systems; c) integrating incomplete state information into stochastic reachability calculations.* We will now give an overview of stochastic reachability, and summarize preliminary work in the areas listed above along with their potential impact on CPS.

Stochastic reachability analysis has been established as a tool for generating probabilistic claims of safety, where rather than considering whether a failure may occur, it gives the *probability* of failure. Its development is mainly in application to stochastic hybrid systems, whose integration of both continuous and discrete co-evolving states makes them an ideal model for CPS. In particular, a discrete time stochastic hybrid system (DTSHS) is a tuple $H = (\mathcal{X}, \mathcal{Q}, \mathcal{U}, T_x, T_q)$ where $\mathcal{X} \subseteq \mathbb{R}^n$ is a set of continuous states, $\mathcal{Q} = \{q^1, q^2, \dots\}$ is a set of discrete modes, $\mathcal{U}$ is the set of possible control inputs, and T_x and T_q are stochastic transition kernels governing continuous and discrete state updates, respectively.

For a predetermined safe set of states K, the probability of a DTSHS remaining within the safe region for a given time horizon N can be determined, and stochastic viable sets (set of all initial states (x_0, q_0) remaining within K up to time N) generated. Given that for a random variable X, the probability $\mathbb{P}[x \in K] = \mathbb{E}[\mathbf{1}_K(x)]$, with $\mathbb{E}$ denoting expected value and indicator function $\mathbf{1}_K(x) = 1$ if $x \in K$ and $\mathbf{1}_K(x) = 0$ otherwise, the maximal probability of staying within region K starting in intial state (x_0, q_0) is written as:

$$\max_{u \in \mathcal{U}} \mathbb{E}\left[\prod_{n=1}^{N} \mathbf{1}_K(x_n, q_n) \middle| x_0, q_0 \right]. \tag{1}$$

Solving the stochastic optimal control problem (1) gives an upper bound to the probability that the hybrid state stays outside of the unsafe region $\overline{K}$ for time steps $n = 1, \dots, N$, and also provides an optimal controller for ensuring safety. As a sequential decision-making problem, (1) can be solved using dynamic programming, and the optimal controller given in the form of a look-up table. Although dynamic programming is known to scale poorly to higher dimensional systems, we have applied this technique to a number of important applications, from human in-the-loop anesthesia delivery to motion planning with 50 moving obstacles, in addition to investigating alternate computational methods.

a) Automated anesthesia delivery with the possibility for human input is an excellent example of a CPS where guarantees of safety are crucial. Consider a pharmokinetic model describing the concentration of the anesthetic Propofol in three compartments of the body (given by $x_i(t)$, $i = 1, 2, 3$). The change in concentration is governed by a difference equation with both a continuous (automated infusion) control input and a discrete bolus dose that may be delivered by the anesthesiologist. The anesthesiologist does not act with certainty, however their actions can be approximately modeled by a Markov chain. Stochastic reachability can then be used to determine the probability that a patient's concentration of Propofol remains within safe tolerance levels for the duration of the surgery (see Fig. 1), despite uncertain input. Our preliminary work assumes an oversimplified model of human decisions, and much more work is needed to develop accurate models of human behavior.

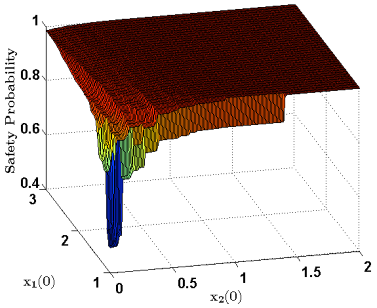

Fig. 1: Viability probabilities for anesthesia delivery with $x_3(0) = 0.01$.

b) Many systems are too large for dynamic programming solutions, hence the need for efficient ways to compute stochastic viable sets. We have begun to collaborate with the Air Force Research Laboratories to develop computational methods for generating stochastic reach-avoid sets for spacecraft rendezvous (a six-dimensional problem). A rendezvous is deemed a success when the approaching craft *avoids* the area outside the line-of-sight cone of the target craft, and *reaches* the target without overshooting or colliding. Fig. 2 was generated using a particle approximation to the reach-avoid probability, and solved as a mixed integer linear program. Improvements in calculation speed and in the incorporation of feedback controller structures are still necessary.

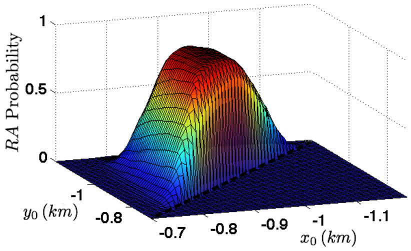

Fig. 2: Reach-avoid probabilities for spacecraft rendezvous with fixed initial velocity.

c) In CPS characterized by interacting autonomous and human agents, an asymmetric and imperfect sharing of information can dramatically alter safety-based controllers. Stochastic reachability can provide a framework that determines how systems will perform under various degrees of information sharing. We have developed a sufficient statistic $\eta_n(\rho, i_n) = \mathbb{E}\left[\mathbf{1}_q(q_n)\mathbf{1}_x(x_n)\prod_{i=1}^{n-1}\mathbf{1}_K(x_i, q_i)\,\middle|\, i_n\right]$, with ρ the initial distribution of (x_0, q_0) and i_n the sequence of observations and control inputs up to time n, for transforming (1) and the associated dynamic programming equations into an equivalent problem of perfect information, when only noisy observations of the discrete state q_n and continuous state x_n are available. We are currently exploring a modified point-based value iteration (PBVI) algorithm to facilitate computation in this much more expensive problem. We demonstrated the method on a one-dimensional temperature regulation problem as proof of concept (see Fig. 3), however much more work needs to be done to extend this to systems of moderate dimension.

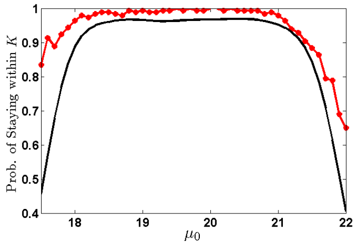

Fig. 3: Viability probabilities for 1-D heater example, calculated using PBVI (black) and simulated using controller produced by PBVI (red) for varying initial temperature distribution.

Stochastic reachability provides a theoretical framework for identifying probabilistic assurances of safety in safety critical, high-risk, or expensive CPS before they are constructed or deployed. Further research into the incorporation of incomplete information will provide the formal foundations for verification and understanding of realistic CPS, and development of efficient computational methods will make these foundations applicable to a broad range of systems, including those with human decision-makers. *The potential impact of stochastic reachability on CPS is evident for practical problems in healthcare and transportation technologies like those introduced here, amongst other safety critical systems. With continued research, stochastic reachability will improve the safety and autonomy of CPS, which is essential for their advancement as the new standard of engineered systems.*