Poster: Inter-Triggering Hybrid Automata: A Formalism for Responsibility-Sensitive Safety

Necmiye Ozay
University of Michigan, Ann Arbor, USA

ABSTRACT

This paper introduces inter-triggering hybrid automata, a formalism to represent multi-agent systems where each agent is represented as a hybrid automaton and agents interact by triggering discrete transitions (jumps and resets) on their "neighboring" agents. Using this formalism, we define responsibility-sensitive safety as respecting one another's invariances while triggering jumps and resets. This allows us to make a formal connection between responsibility and robust controlled invariant sets for individual agents, therefore leading to a compositional verification framework for the safety of the overall multi-agent system. We discuss several advantages of this viewpoint and illustrate it on a highway driving example.

ACM Reference Format:
Necmiye Ozay. 2020. Poster: Inter-Triggering Hybrid Automata: A Formalism for Responsibility-Sensitive Safety. In *23rd ACM International Conference on Hybrid Systems: Computation and Control (HSCC '20), April 22–24, 2020, Sydney, NSW, Australia.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3365365.3383470

1 INTRODUCTION

Verification of safety of multi-agent systems, especially when some of the agents are autonomous and some are human-controlled, is an important challenge. Prime examples of such systems include autonomous mobile robots in urban environments (e.g., sidewalks, shopping malls, museums) or autonomous cars in traffic. The challenges arise due to a number of factors. For instance, as the number of agents can be large, verification methods need to be scalable. Another challenge is the trade-off between safety and conservativeness. When designing an autonomous agent to operate among non-autonomous agents, assuming all the non-autonomous agents to be adversarial leads to unnecessary conservativeness.

In this paper, we introduce *Inter-triggering Hybrid Automata*, a collection of hybrid automata that interact by triggering the jumps an resets (i.e., discrete transitions) on one another. Each hybrid automaton is modeling an individual agent, and discrete transitions are due to interactions. Then, we define responsibility as respecting one another's invariances while triggering jumps and resets. This allows us to separate the individual invariant computations from reasoning about the behavior of the collection, thus leading to a compositional and modular framework for synthesis and verification. We show an application of this framework on highway driving where triggering is due to lane changes. Several advantages of the proposed framework are discussed.

HSCC '20, April 22–24, 2020, Sydney, NSW, Australia

ACM ISBN 978-1-4503-7018-9/20/04.
https://doi.org/10.1145/3365365.3383470

2 INTER-TRIGGERING HYBRID AUTOMATA

Let $\mathbb{T}$ denote the time domain, which can be discrete-time ($\mathbb{T} = \mathbb{N}$) or continuous-time ($\mathbb{T} = \mathbb{R}_{\geq 0}$).

Definition 1. *An* inter-triggering hybrid automata *is a collection* $\{\mathcal{H}_i, N_i\}_{i\in\mathcal{I}}$ *of parametrized hybrid systems, where* $\mathcal{I}$ *is a countable index set, each* $\mathcal{H}_i$ *is a hybrid system of the form* $\mathcal{H}_i = \langle \Gamma_i, X_i, R_i, F_i\rangle$ *representing the agent* i*, where:*

- $\Gamma_i = (Q_i, E_i)$ *is a directed graph where* Q_i *is a finite set of discrete-states or modes and* $E_i \subseteq Q_i \times Q_i$ *is a set of edges,*
- $X_i = \{X_{i,q}\}_{q\in Q_i}$ *is a set of continuous domains, where* $X_{i,q} \subseteq \mathbb{R}^{n_{i,q}}$,
- $R_i = \{R_{i,e}\}_{e\in E_i}$ *is a set of reset maps with* $R_{i,(q,q')} : X_{i,q} \to 2^{X_{i,q'}}$ *for* $e = (q, q')$,
- $F_i = \{f_i^q\}_{q\in Q_i}$ *is a set of vector fields, where* $f_i^q : X_{i,q}\times U_i\times D_i \to \mathbb{R}^{n_{i,q}}$ *governs the continuous-state update equation on* $X_{i,q}$*, and* U_i *and* D_i *are continuous input and disturbance sets, respectively, and each* $N_i : \mathbb{T} \to 2^{\mathcal{I}}$ *represents the time-varying set of reset-triggering neighbors for agent* $\mathcal{H}_i$.

From the perspective of agent i, the set $N_i(t)$ of agents triggering the discrete transitions and the outcome of resets can be seen as external (adversarial) signals or it can be state-dependent. Moreover, if $i \in N_i(t)$, the agent can self-trigger a discrete transition. If multiple agents instantaneously try to trigger a discrete transitions on agent i, there is a priority order that picks the appropriate reset outcome. We do not specify these in the generic definition above but, they can be specified in the context of the application under consideration. In particular, we will specify them in the context of car following with lane changes problem presented next.

2.1 Highway Driving Example

Consider a car following scenario on the highway, where the ego vehicle and its immediate interaction with a lead vehicle can be represented with a discrete-time hybrid automaton $\mathcal{H}_i$ as in Fig. 1. The continuous state consists of the velocity v_e of the ego car, the lead car velocity v_L, and the longitudinal headway h_L with respect to the lead car, the input is the ego car acceleration $a_e \in [\underline{a}_e, \overline{a}_e]$, the acceleration $a_L \in [\underline{a}_L, \overline{a}_L]$ of the lead car is treated as a disturbance, and Δt is the sampling time. For a state variable x, x^+ denotes its value in the next time step. The

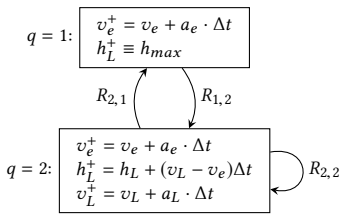

Figure 1: A hybrid system representation of a car following scenario (adapted from [2]).

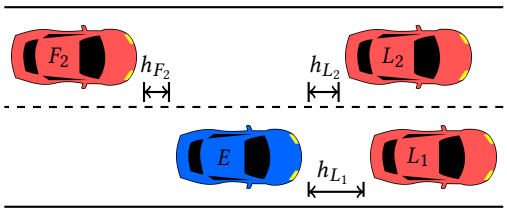

Figure 2: A collection of vehicles on the highway.

mode $q = 1$ corresponds to the case where there is no lead car in front of the ego vehicle and $q = 2$ is when there is a lead vehicle, with the discrete transitions $R_{2,1}$, $R_{1,2}$, and $R_{2,2}$, capturing the lead car leaving the lane with no other lead car present or the ego car merging to a lane where there is no lead car, a lead car cutting in front of the ego car or ego car merging to a lane where there is a lead car while initially there is no lead car, and a different lead car cutting in front of the ego car or ego car merging to a lane where there is a different lead car, respectively.

Now consider a collection $\mathcal{I} = \{E, L_1, L_2, F_2\}$ as in Fig. 2. Each vehicle $i \in \mathcal{I}$ in this collection can be represented with an hybrid automaton $\mathcal{H}_i$ as in Fig. 1 in their local coordinates. Moreover, they can trigger discrete transitions on one another by performing lane change actions (for simplicity, assumed to be instantaneous). Let Fig. 2 be a snapshot at time $k \in \mathbb{N}$, then we have $N_E(k) = \{E, L_1, L_2\}$, $N_{L_1}(k) = \emptyset$, $N_{L_2}(k) = \{L_1, L_2\}$, and $N_{F_2}(k) = \{L_2, F_2, E\}$ for reset-triggering neighbors for each agent. Therefore, overall collection can be represented as inter-triggering hybrid automata.

As for specifications, we focus on safety specifications. These consist of, for each vehicle, to maintain a safe distance from the lead vehicle (when in mode $q = 2$) and to obey the speed limits in both modes. Therefore, we can associate to each mode a safe set $\mathcal{X}^1_{safe} = \{(v_e, h_L) \mid \underline{v}_e \le v_e \le \overline{v}_e\}$ and $\mathcal{X}^2_{safe} = \{(v_e, h_L, v_L) \mid \underline{v}_e \le v_e \le \overline{v}_e,\ h_L \ge \underline{h}_L\}$. Safety amounts to guaranteeing that the states remain in these sets indefinitely.

3 RESPONSIBILITY-SENSITIVE SAFETY VIA INVARIANCE

If we model the highway driving using an inter-triggering hybrid automata, from the perspective of a single agent, if the other agents are allowed to adversarially trigger the resets, remaining invariantly inside $\mathcal{X}^1_{safe}$ or $\mathcal{X}^2_{safe}$ is not possible. For example, if a very slow car suddenly cuts right in front of a safety supervisor/controller equipped car on the highway, no supervisor/controller can prevent a crash. On the other hand, drivers on the road (agents in our hybrid collection) do not act adversarially but they are expected to behave "responsibly". Inspired by the work on responsibility-sensitive safety [4, 5], we use the introduced hybrid system model to formalize the correct behavior of a supervisor in interactive highway settings.

We first note that, if we ignore the possibility of a discrete transition (i.e., lane-change), we can compute an invariant set C^i_{inv,q_i} for each agent i in each mode q_i inside the respective safe set.

Definition 2. *Let agent i with model $\mathcal{H}_i$ trigger a discrete transition on agents J. Then, we have the following two rules for safety:*

- **(self-safety)** *If $i \in J$, the discrete-transition from q_i to q_i' triggered must be such that $x_i \in C^i_{inv,q_i}$ implies $x_i' \in C^i_{inv,q_i'}$. All agents remain in their respective invariants when there is no discrete-transition.*
- **(responsibility)** *For $j \in J$, the discrete-transition from q_j to q_j' agent i triggers must be such that $x_j \in C^j_{inv,q_j}$ implies $x_j' \in C^j_{inv,q_j'}$.*

It can be shown that if all agents follow the local rules in Def. 2, each agent is guaranteed to remain in their safe sets indefinitely.

Discussion: Inter-triggering hybrid automata naturally addresses some of the challenges in verification of responsibility sensitive safety mentioned in [3]. In particular, it has the following desirable properties:

- *From verification to synthesis:* Maximal invariants, once computed for each agent, can be used to verify the responsibility sensitive safety of the collection. Moreover, if we replace invariant sets for given policies with maximal controlled invariant sets (CIS), a compositional synthesis problem can be posed for the multi-agent system.
- *Incorporating agent-to-agent communication:* When we do not control all the agents, we might not know the invariants of others (therefore what our responsibility is). V2V communication can be used between neighbors to guarantee responsibilities are not violated.
- *Incorporating risk in planning and control:* If communication is not possible (e.g., V2V technology or the knowledge of invariants are not available for some vehicles) and if we are in the synthesis setting, we need to make assumptions on the behavior of surrounding vehicles (disturbance bounds, vehicle models) both to compute our invariants and to predict others'. This gives two tuning knobs where one can assume aggressiveness and agility of the others, leading to a smaller robust CIS for the self and larger robust invariant predictions for the others. These assumptions can be used to tune conservativeness and risk.
- *Handling imperfections in sensing and actuation:* There is recent progress on computing invariant sets for systems with delays or in imperfect measurement settings. Since our framework separates invariant computation from verification of responsibility sensitive safety, such imperfections can be naturally handled by leveraging results from invariant computation.

Future work involves investigating the connection of our framework to hybrid I/O automata [1] and thorough theoretical analysis.

Acknowledgments: The author thanks Kwesi Rutledge, Glen Chou, and Michael Arwashan for insightful discussions and for their help with the example. This work is supported by NSF grant no. 1918123.

REFERENCES

[1] N. Lynch, R. Segala, and F. Vaandrager. Hybrid i/o automata. *Information and computation*, 185(1):105–157, 2003.
[2] P. Nilsson, O. Hussien, A. Balkan, Y. Chen, A. D. Ames, J. W. Grizzle, N. Ozay, H. Peng, and P. Tabuada. Correct-by-construction adaptive cruise control: Two approaches. *IEEE Trans. on Control Syst. Technol.*, 24(4):1294–1307, 2016.
[3] N. Roohi, R. Kaur, J. Weimer, O. Sokolsky, and I. Lee. Self-driving vehicle verification towards a benchmark. *arXiv preprint arXiv:1806.08810*, 2018.
[4] S. Shalev-Shwartz, S. Shammah, and A. Shashua. On a formal model of safe and scalable self-driving cars. *arXiv preprint 1708.06374*, 2017.
[5] S. Vaskov, S. Kousik, H. Larson, F. Bu, J. Ward, S. Worrall, M. Johnson-Roberson, and R. Vasudevan. Towards provably not-at-fault control of autonomous robots in arbitrary dynamic environments. *arXiv preprint arXiv:1902.02851*, 2019.