ARCH-COMP Report: Preliminary Results on the Falsification Benchmarks

Adel Dokhanchi, Shakiba Yaghoubi, Bardh Hoxha, and Georgios Fainekos

Arizona State Univeristy, School of Computing, Informatics and Decision Systems Engineering,
Tempe, AZ, U.S.A.
{adokhanc, syaghoub, bhoxha, fainekos}@asu.edu

1 Introduction

This report presents some preliminary results for the 2017 friendly competition of the ARCH workshop[1] for falsification of temporal logic specificaitons over Cyber-Physical Systems. The benchmarks are available on the ARCH website (cps-vo.org/group/ARCH). In this report, we present results on a powertrain model developed by Toyota Technical Center which contains a complex automatic air-fuel control subsystem [6].

2 Falsification Tool: S-TaLiRo

S-TaLiRo [2] is a Matlab toolbox that searches for system behaviors that falsify (do not satisfy) specifications presented in Signal Temporal Logic (STL) [7]. It can analyze arbitrary Simulink models or user defined functions that model the system. S-TaLiRo performs automated randomized testing based on stochastic optimization techniques. Among the advantages of the toolbox is the seamless integration inside the Matlab environment, which is widely used in the industry for model-based development systems. For an overview of the S-TaLiRo functionality see [5].

3 Benchmark Results

Our experiments were conducted on a 64-bit Intel Xeon CPU (2.5GHz) with 64-GB RAM and Windows Server 2012. We used MATLAB 2015a to run the falsification toolbox S-TaLiRo [2]. For our experiments, we used the following stochastic optimization methods: Simulated Annealing (SA) [3], Cross-Entropy (CE) optimization [8] and Uniform Random (UR) sampling. We remark that all the experiments were performed with the default parameters for each optimization method. It would be expected that further improvements can be achieved by tuning the performance of the optimization algorithms for each benchmark problem. All the benchmark problems are available with the S-TaLiRo distribution [2] or from the ARCH workshop repository [1].

[1] Workshop on Applied Verification for Continuous and Hybrid Systems (ARCH), cps-vo.org/group/ARCH

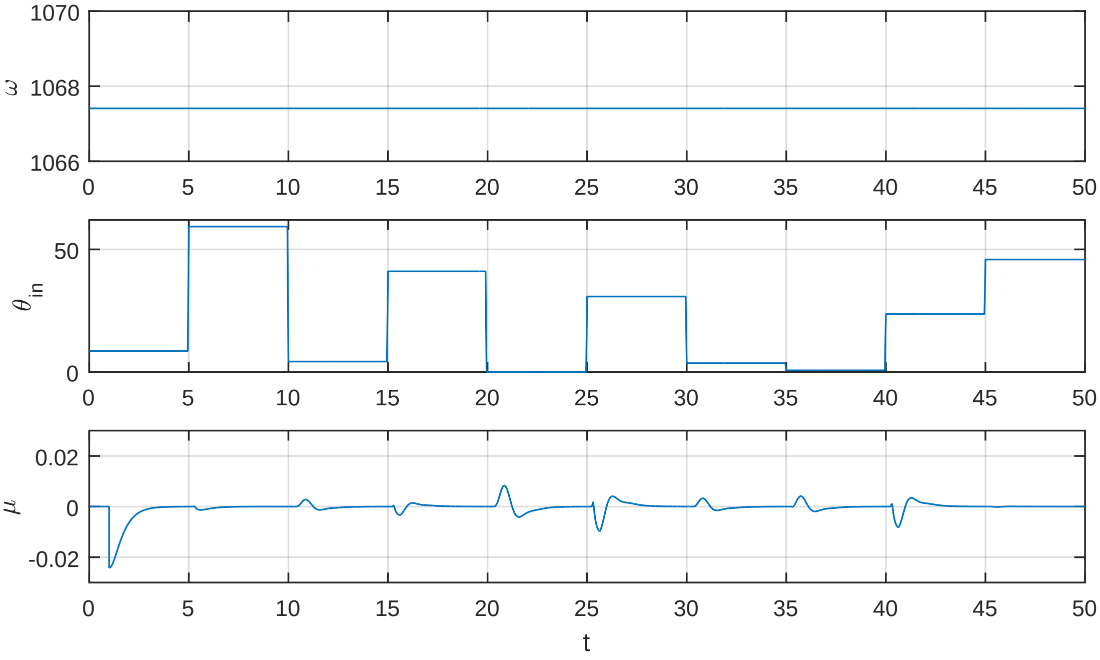

Figure 1: A sample piecewise constant falsifying input signal and the corresponding trajectory of the powertrain system for ϕ_{PB}. The formula is falsified at time t=20.785, and the robustness value is -6.12×10^{-5}. The input focuses on antecedent falsification up to t=25.

3.1 Powertrain Control

The Powertrain Control benchmark presented in this report was first introduced in [6]. The benchmark provides a high complexity model of an automatic air-fuel control system. It consists of an air-fuel controller and a mean-value engine model. The closed loop system takes two exogenous inputs: the throttle angle θ_{in} and, the engine speed ω. It has 3 continuous-valued states associated with the controller and 5 continuous-valued states associated with the plant. In addition, we have the states which are introduced by the variable delay.

The controller has 4 modes of operation: "Startup", "Normal", "Power" and "Fault". Depending on the operation mode, the system should satisfy different requirements. We used a slightly modified version of the requirements presented in Eq. (27) of the paper by Jin et al. [6]. This following specification needs to be satisfied when the system is in the "Normal" mode:

$$\phi_{PB} = \Box_{(\tau_s,T)}((rise(a) \vee fall(a)) \rightarrow \Box_{(\eta,\zeta)}(|\mu| < \beta))$$

where $a = 40$, $rise(a) \equiv (\theta_{in} \leq 8.8°) \wedge \Diamond_{(0,\epsilon)}(\theta_{in} \geq a)$ and $fall(a)$ is defined similarly for an small enough ϵ, $\tau_s = 11$ is the necessary time for the system to enter the "Normal" mode from the "Startup" mode, $T = 50$ is the simulation time, $\eta = 1$ is the settling time required after $rise$ or $fall$ happens, $\zeta = 5$ is the end of the current time interval in which the input is kept constant, μ is the normalized error signal that indicates the error in the value of the state Air/Flow ratio from a reference value.

The formula states that whenever event $rise$ or $fall$ happens (the antecedent, which is over the input signal), μ should remain in the specified bound after the settling time η, and before other changes are made to the input. The antecedent of the formula is over the input signals of the system. In this report, the acceptable error bound β is reduced to 0.008 to make falsification feasible. Note that abrupt changes in the value of inputs are acceptable and necessary here to satisfy the antecedent but, frequent changes in the input are not. As a matter of fact, decreasing the frequency of the changes make the problem less interesting.

Table 1: General Falsification

Optim.	Fals	Min n.tests	Max n.tests	Avg n.tests	Min Rob.	Max Rob.	Avg Rob.
UR	7/50	18	93	52	1.7×10^{-5}	0.0035	8.81×10^{-4}
SA	9/50	13	83	50	3.54×10^{-5}	0.0042	0.0012
P-SA	4/50	34	80	55	7.41×10^{-6}	0.0051	0.0016

Table 2: Vacuity Aware Falsification

Optim.	Fals.	Min n.tests	Max n.tests	Avg n.tests	Min Rob.	Max Rob.	Avg Rob.
UR	9/50	12	96	63	3.4×10^{-6}	0.003	0.00086
SA	29/50	7	95	39	2.38×10^{-6}	0.0043	0.0013

We compare results for two falsification algorithms. One is a general falsification algorithm, where the optimizer minimizes the robustness value with respect to a STL specification.This is the standard method used in S-Taliro. The second one is Vacuity Aware Falsification (VAF) similar to [4]. In VAF, for reactive specifications, as a first step in falsification, we attempt to satisfy the antecedent and then falsify the specification. This functionality has not been released yet to the public S-Taliro repository. Since the antecedent can be falsified at any time after τ_s, in our experiments we tried to falsify it in a fraction of T ($T/2$ here) so that we have enough time in the future to falsify the whole formula.

We used 50 runs for each algorithm with 100 tests. The experimental results and a sample falsifying input and trajectory are shown in Tables 2 and 1 and Fig. 1, respectively. "Min n.tests" show the minimum number of tests in the case of falsification, while "Min Rob." indicate the minimum best robustness values achieved for the cases without falsification, this gives an idea on how close these cases were to falsification. Using VAF, we achieve a slight improvement on the performance of the algorithm. This is due to the fact that the challenge in this falsification benchmark is mainly related to the consequent rather than the antecedent. Generally, if we heuristically force the antecedent to occur in the first half of the trace, then we observe a considerable increase in the number of falsifications. However, we cannot claim that because we enforce the antecedent to occur earlier, there is more time to search for the consequent. In this benchmark example, the consequent must occur within 5 time units of the antecedent being activated. Therefore, we believe that there is space for improvement in the falsification rate even for pure black-box methods and that this is a challenging benchmark which can drive forward the competition in the falsification category of the ARCH workshop.

We also designed another experiment in which the antecedent is always satisfied when we are sampling for new input signals. Since the antecedent is satisfied whenever $rise$ or $fall$ happens, we used a single pulse as the input signal (shown in Fig 2). The search space in S-TaLiRo is over the times t_1 and t_2 and the signal values x_1,x_2 and x_3 such that the antecedent is always

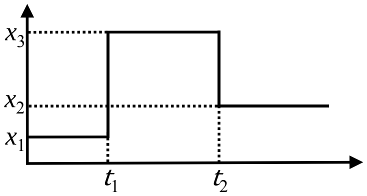

Figure 2: Pulse input to satisfy antecedent of ϕ_{PB}.

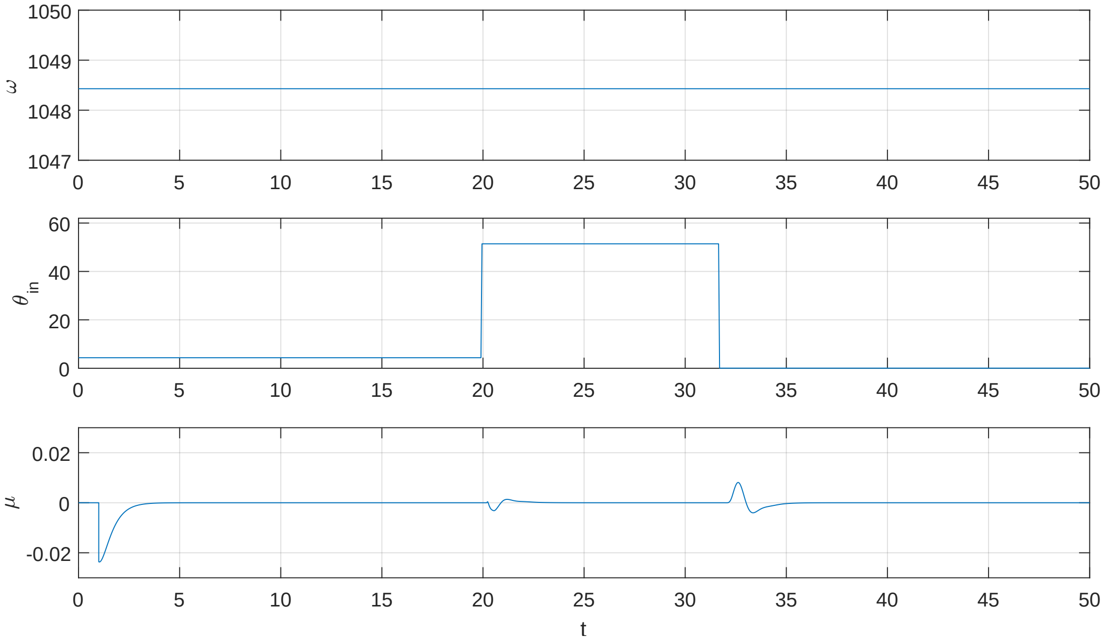

Figure 3: A sample falsifying pulse input signal and the corresponding trajectory of the powertrain system for ϕ_{PB}. The formula is falsified at time t=32.6425, and the robustness value is -4.083×10^{-5}.

satisfied ($t_2 < t_1 + 5$, $x_1, x_3 < 8.8$ and $x_2 > a$). Note that we can also try the case for the inverse pulse in which $x_1, x_3 > a$ and $x_2 < 8.8$. However, allowing the input to have either the first set of constraints or the second set of constraints will make the problem non-convex. Although in this search, the input is constrained to satisfy the antecedent, only in 4 out of 50 runs S-TaLiRo has been able to falsify ϕ_{PB}. This result is shown in Table 1 under the name "P-SA". A falsifying input is shown in Fig 3.

4 Conclusions

We have presented some preliminary base results for the falsification competition of the ARCH workshop. The results indicate that black box search based test generation methods do not perform much better than random sampling on this challenging benchmark. On the other hand, utilizing some information on the structure of the specification can help in doubling the rate of falsifications. We hope that this is viewed as a motivation that there is space for improvement in black box or gray box falsification methods and a competition on falsification.

References

[1] Applied Verification for Continuous and Hybrid Systems (ARCH) http://cps-vo.org/group/ARCH.

[2] S-TaLiRo : https://sites.google.com/a/asu.edu/s-taliro/.

[3] H. Abbas, G. Fainekos, S. Sankaranarayanan, F. Ivančić, and A. Gupta. Probabilistic temporal logic falsification of cyber-physical systems. *ACM Trans. Embed. Comput. Syst.*, 12(2s):95:1–95:30, May 2013.

[4] T. Akazaki. Falsification of conditional safety properties for cyber-physical systems with gaussian process regression. In *Runtime Verification - 16th International Conference, RV 2016, Madrid, Spain, September 23-30, 2016, Proceedings*, pages 439–446, 2016.

[5] B. Hoxha, H. Bach, H. Abbas, A. Dokhanchi, Y. Kobayashi, and G. Fainekos. Towards formal specification visualization for testing and monitoring of cyber-physical systems. In *Int. Workshop on Design and Implementation of Formal Tools and Systems*. October 2014.

[6] X. Jin, J. V. Deshmukh, J. Kapinski, K. Ueda, and K. Butts. Powertrain control verification benchmark. In *Proceedings of the 17th international conference on Hybrid systems: computation and control*, pages 253–262. ACM, 2014.

[7] O. Maler and D. Nickovic. Monitoring temporal properties of continuous signals. In *Proceedings of FORMATS-FTRTFT*, volume 3253 of *LNCS*, pages 152–166, 2004.

[8] S. Sankaranarayanan and G. Fainekos. Falsification of temporal properties of hybrid systems using the cross-entropy method. In *Proceedings of the 15th ACM International Conference on Hybrid Systems: Computation and Control*, HSCC '12, pages 125–134, New York, NY, USA, 2012. ACM.