

# TLTk: A Toolbox for Parallel Robustness Computation of Temporal Logic Specifications

Joseph Cralley <sup>1</sup>, Ourania Spantidi <sup>1</sup>, Bardh Hoxha <sup>2</sup>, and Georgios Fainekos <sup>3</sup>

<sup>1</sup> Southern Illinois University, Carbondale IL 62901

{jkolecr, ourania.spantidi}@siu.edu

<sup>2</sup> Toyota Research Institute North America, Ann Arbor, MI 48105

bardh.hoxha@toyota.com

<sup>3</sup> Arizona State University, Tempe, AZ 85281

fainekos@asu.edu

**Abstract.** This paper presents the Temporal Logic Toolkit (TLTk), a modular falsification tool for signal temporal logic specifications developed in Python and C. At the core of the tool, an algorithm for robustness computation is utilized that supports multi-threaded CPU/GPU computation. The tool enables memory-efficient, parallel, robustness computation of system traces. In addition, the python implementation enables the addition and modification of temporal operators for application-specific scenarios. The performance of the tool is evaluated against state-of-the-art robustness computation engines DP-TALiRO and BREACH on a number of benchmark problems.

**Keywords:** Testing · Temporal Logic · Robustness.

## 1 Introduction

The theory of robustness of temporal logics [13] has been utilized in a wide-array of problems, from testing and verification of Cyber-Physical Systems (CPS) to monitoring and planning for autonomous systems [13,10,29,15]. It enables the formulation of the falsification problem [25,22], i.e. the problem of finding system behaviors that do not meet system requirements, as a non-convex, non-linear optimization problem. The falsification process uses a notion of robustness to indicate how well a trajectory satisfies a requirement. This robustness estimate is defined using quantitative semantics of temporal logics such as STL (see [4] for an overview). The robustness indicates by how much a trajectory may be perturbed without changing the Boolean truth value of the specification. In the falsification process, the robustness is used to guide the optimization function to search for regions in the set of inputs and initial conditions of the system in which falsification is more likely.

Falsification, and the related problem of parameter mining [19,21,18], have been used successfully for testing industrial-size CPS. Both of these methods have been successfully used in a wide array of applications, from medical device testing [6], engine control testing [20,18,17], Unmanned Aerial Vehicles (UAV)

scenario generation [29], to Automated Driving Systems [28,16]. In each optimization loop in the falsification process, the two main computational elements are the system simulator and the robustness computation engine. To improve this process, we introduce TLTK<sup>1</sup>, a Python/C toolkit for requirements-based testing of CPS. TLTK is developed with the goal of optimizing the robustness computation engine as much as possible. At the core of the tool, a robustness computation engine that supports multi-threaded CPU and GPU computations is utilized. The memory-efficient algorithm enables robustness computations of large system traces. In addition, the robustness algorithm written in Python/C allows for easy modification/addition of temporal logic operators for application-specific implementations. This is particularly useful in areas such as planning for robotic applications since notions of robustness are usually application-specific.

TLTK supports falsification for STL specifications using only open-source software. Also, we provide a repository through the OS-virtualization engine Docker that allows easy integration with other tools or deployment in large-scale cloud systems like Amazon AWS, Google Cloud or Microsoft Azure. TLTK has been successfully utilized with several benchmark problems from the CPS community. The performance of the tool in comparison to state-of-the-art tools BREACH [9] and DP-TALiRO [11] is presented.

## 2 Overview and Features

TLTK is an object-oriented toolbox developed in `python3` (front-end) and `C` (back-end). An overview of the tool is presented in Figure 1. The toolbox has the following core modules:

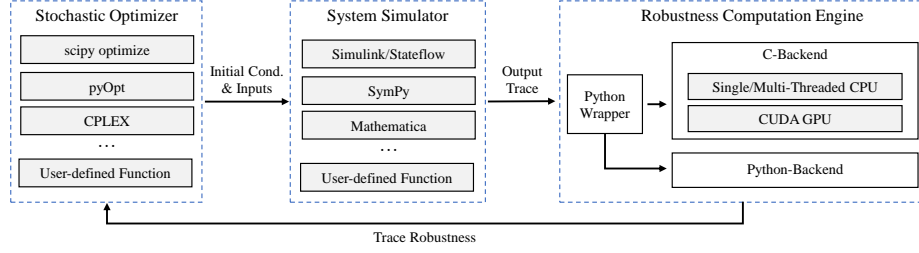
1) The **Stochastic Optimizer** module is developed in python and is utilized to generate candidate initial conditions and input signals for the system [1]. Our implementation utilizes global optimization algorithms provided by the SciPy library<sup>2</sup> such as Dual Annealing [30], as well as local optimization algorithms such as Nelder-Mead [24] for refinement. In addition, due to the modular architecture of the tool, the user may develop their own or utilize any other optimization libraries in python to conduct the search.

2) The **System Simulator** module may be a standalone function, an interface for Matlab/Simulink, or other simulators that support python-integration such as SymPy and Mathematica.

3) The **Robustness Computation Engine** module utilizes our C back-end implementation for fast robustness computation. The module utilizes multi-threaded CPU/GPU processing to compute the robustness of a trace with respect to a specification in parallel. A python implementation is also available. Although much slower than the C implementation, the python implementation

<sup>1</sup> The source code for TLTK is publicly available through the `git` repository: <https://bitbucket.org/versyslab/tltk/>. Docker image: <https://hub.docker.com/r/bardhh/tltk>. Python package through PyPi: <https://pypi.org/project/tltk-mtl/>. User Guide: <http://www.bhoxha.com/tltk>.

<sup>2</sup> SciPy Optimize: <https://docs.scipy.org/doc/scipy/reference/optimize.html>



**Fig. 1.** An overview of TLTK and its major components.

is developed to make the tool more accessible and also so that modifications to the robustness algorithm can be made easily for utilization in application-specific case studies and even topics such as planning. For example, in [23,15], the authors consider a smooth cumulative robustness, which modifies the semantics of temporal operators to include smoothing functions in order to make it more suitable for planning problems. For specifications with higher dimensional predicates, an optimization problem needs to be solved in order to compute the distance between a point in a trajectory to an unsafe set. Therefore, the back-end robustness calculation module additionally calls the quadratic program solver QuadProg<sup>3</sup>. To setup and run the tool, several options are available:

- *Building from Source* (Linux). This option provides the best performance. However, it is restricted to the Linux OS since we are using OpenMP<sup>4</sup> for parallel computation. The user needs to follow a set of commands provided in the User Guide to install software dependencies and compile the programs.
- *Running through Docker* (Linux, Windows, Mac). Docker enables a single command setup for the tool and all the required dependencies. Currently, GPU functionality is accessible only on Linux hosts<sup>5</sup>. The TLTK docker image can be pulled using the following command: `docker pull bardhh/tltk`.
- *Python API* (Linux). In addition to the previous methods, TLTK is available as a python package and can be installed through the pip3 Python package installer using the following command: `pip3 install tlkt_mtl`. Once the package is installed, it can be imported in any python3 script and used for robustness computation.

### 3 Robustness Computation

First, we review the quantitative semantics of STL specifications that enable us to define the robustness estimate. After that we propose a parallel algorithm for computing the robustness estimate efficiently.

<sup>3</sup> QuadProg: <https://github.com/rmcgibbo/quadprog>

<sup>4</sup> OpenMP: <https://www.openmp.org/>

<sup>5</sup> Nvidia has announced that support for this functionality in Windows OS is under development. <https://devblogs.nvidia.com/announcing-cuda-on-windows-subsystem-for-linux-2/>

### 3.1 Quantitative Semantics of STL Specifications

STL is an extension of LTL that enables reasoning over real-time properties of CPS. The syntax of STL is defined as follows:

$$\varphi ::= \top \mid p(x) \geq 0 \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid \varphi_1 \mathcal{U}_I \varphi_2$$

where  $\top$  is *true* and  $I$  is a nonsingular interval of positive reals. The eventually operator is defined as  $\Diamond\phi \equiv \top \mathcal{U}_I \phi$  and the always operator is defined as  $\Box_I \phi \equiv \neg\Diamond_I \neg\phi$ . In order to define the quantitative semantics of STL over arbitrary predicates  $p(x) \geq 0$ , we use a metric  $d$  [13] to define the distance of a point  $x \in X$  from a set  $S \subseteq X$  as follows:

**Definition 1 (Signed Distance)** *Let  $x \in X$  be a point,  $S \subseteq X$  be a set and  $d$  be a metric on  $X$ . Then, we define the Signed Distance from  $x$  to  $S$  to be*

$$\text{Dist}_d(x, S) := \begin{cases} -\inf\{d(x, y) \mid y \in S\} & \text{if } x \notin S \\ \inf\{d(x, y) \mid y \in X \setminus S\} & \text{if } x \in S \end{cases}$$

The signed distance returns positive values when  $x$  is inside set  $S$  and negative values when  $x$  is outside of set  $S$ .

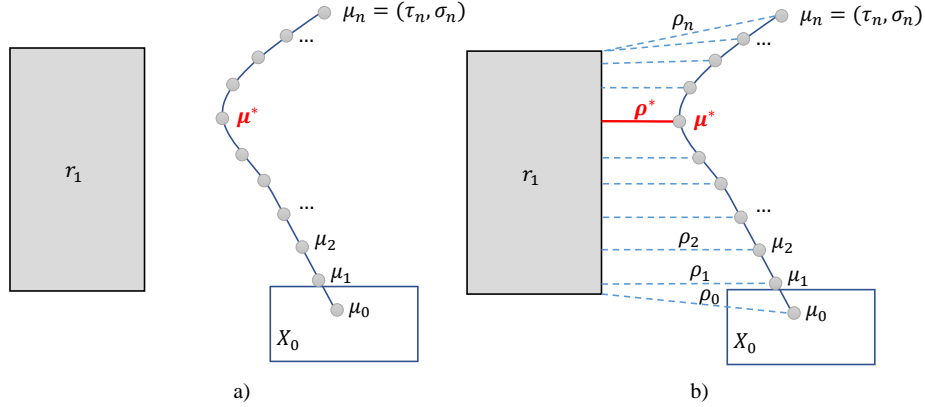
Given a signal  $x$  and an STL specification  $\varphi$ , the quantitative semantics of STL enable us to obtain a robustness degree  $\rho$  that indicates how far the signal is from satisfying or violating the specification starting from a time instance  $t$ . Formally, the robustness of STL specifications is defined as follows:

$$\begin{aligned} \rho(\top, x, t) &:= \infty \\ \rho(p(x) \geq 0, x, t) &:= \text{Dist}_d(x(t), \{x \mid p(x) \geq 0\}) \\ \rho(\neg\varphi, x, t) &:= -\rho(\varphi, x, t) \\ \rho(\varphi_1 \wedge \varphi_2, x, t) &:= \min(\rho(\varphi_1, x, t), \rho(\varphi_2, x, t)) \\ \rho(\varphi_1 \mathcal{U}_{[a,b]} \varphi_2, x, t) &:= \sup_{t' \in [t+a, t+b]} \min\left(\rho(\varphi_2, x, t'), \inf_{t'' \in [t, t']} \rho(\varphi_1, x, t'')\right) \end{aligned}$$

### 3.2 Parallel Robustness Computation Algorithm in TLTK

The parallel robustness computation engine builds on the sliding window / dynamic programming algorithm developed in [31,9]. A tabular representation of subformulas of the specification (rows) and the trace samples (columns) of the signal is utilized. In the rows that contain only atomic predicates, quadratic programming is utilized to compute the cells since robustness requires computations of distances between a point and a set [12]. Since the sets defined by the predicates are strictly convex, we can use quadratic programming with low time complexity. This step is simplified when the predicates define a halfspace in  $\mathbb{R}$ . In this case, the distance may be computed analytically. Following the semantics of the robustness of temporal logic operators, the table can be dynamically collapsed to return a single robustness value for the entire trace.

Consider the specification  $\phi = \Box \neg r_1$ . The specification states that region  $r_1$  should not be reached. In Fig. 2, an example trace and an illustration of the robustness computation process is presented. The robustness value of the trace with respect to the specification should return the minimum distance  $\rho^*$  between the sample point  $\mu^*$  and the unsafe set  $r_1$ . To compute this, the distance of each sample  $\mu_0, \mu_1, \dots, \mu_i, \dots, \mu_n$  to the unsafe set is computed (see Fig. 2 (b)). For each sample, a quadratic program is solved to return the distance to the unsafe set.



**Fig. 2.** An example system trajectory is presented. The specification on the system  $\phi = \Box \neg r_1$  states that the trajectory should not enter set  $r_1$ . For each sample point  $\mu_i$  the minimum distance  $\rho_i$  to the unsafe set  $r_1$  is computed in parallel. The robustness of the trace is  $\rho^*$ , which indicates how much disturbance the trace can tolerate before it does not satisfy the requirement  $\phi = \Box \neg r_1$  any more.

The worst-case time complexity of the robustness computation algorithm is  $O(|\phi||\tau|c)$ , where  $|\phi|$  is the length of the formula,  $|\tau|$  is the number of samples, and  $c$  is the number of samples in the longest interval indicated by the timing constraints. We note that each temporal operator has a time interval associated with it and that the sampling rate is not necessarily constant. Even though the worst-case time complexity is the same as in the algorithm presented in [31], the modifications presented in this section significantly improve the running-time of the algorithm. Specifically, performance improvements are achieved due to parallel computation and more efficient memory management.

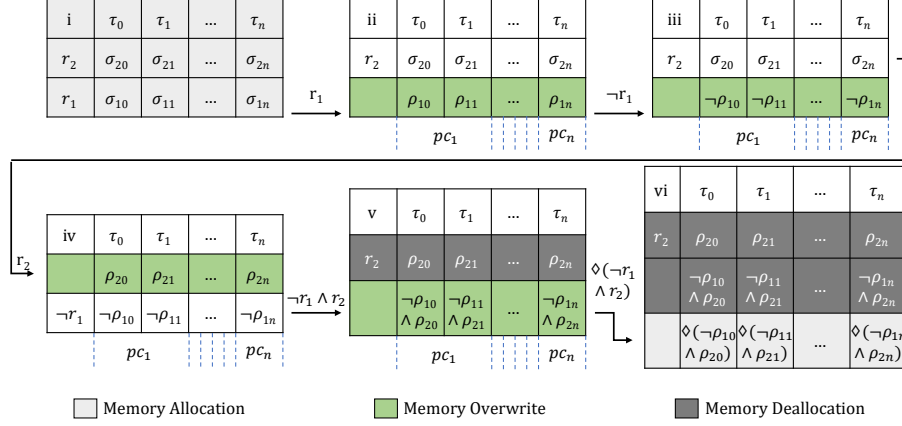
**Parallel Computation:** (i) Given a discrete output trace of the system composed of  $n$  time-value pairs, the robustness computation requires that for each predicate in the formula, the robustness should be computed  $n$  times. This process is parallelized in TLTK with multi-threaded CPU/GPU support. The dynamic programming algorithm generates a table with entries for each subformula and predicate. The bottom rows of the table are reserved for predicates and each distance to the unsafe set is computed in parallel. (ii) Once this is com-

pleted, the tree of subformulas is traversed in parallel to return the robustness. Furthermore, for rows with temporal operators that contain timing intervals, the computation of each cell in the table may be computed independently from the results in the adjacent cells, and therefore the entire row may be computed in parallel. The computation of  $n$  cells is broken up in  $n/c$  groups based on the number of threads  $c$  available. For each thread, a consecutive number of cells is computed. In this case, the sliding window algorithm utilizes the results from the adjacent cell to the right to compute the current cell. This reduces the number of operations within a thread by limiting the min/max operations to only new data in the window.

**Average Running-time:** For time bounded formulas, two improvements are made. (i) A modified binary search algorithm is developed that determines the indices that correspond to the time bounds. For every iteration of the algorithm, the indices are approximated based on the previous indices and then verified. In addition, as the formula is evaluated, the time bound under consideration is restricted to the remaining trace. (ii) For time bounded formulas, there is a sliding window reflecting the time bound of the formula. Since robustness computation is a sequence of min/max operations and most of the data in the current sliding window overlap with the previous window, we only need to consider the changes between the sliding windows to calculate the min/max of the current window.

**Memory Management:** By dynamically allocating and deallocating memory based on the structure of the formula, TLTK operates with a significantly smaller memory footprint. Initially, a formula  $\varphi$  is decomposed into a tree structure. For example, for  $\varphi = \Diamond(\neg r_1 \wedge r_2)$ , memory is allocated for predicates  $r_1$  and  $r_2$  and robustness is computed for the two predicates. Instead of allocating a new row for  $\neg r_1$ , an existing row where  $r_1$  was stored is overwritten with the new values. After, the memory allocated for  $r_2$  is utilized to store the robustness with respect to  $\neg r_1 \wedge r_2$ . The only additional row added is for the final expression of the eventually operator. This process is illustrated in Fig. 3. The worst case space complexity for the robustness computation is  $\mathcal{O}(\beta \times |\tau|)$ , where  $\beta$  is the number of predicates in the formula and  $|\tau|$  is the number of timestamps. A row is preallocated for each predicate in the formula. We note that if a predicate is repeated in the formula, a memory row is allocated for each repetition. For any number of temporal operators, only one additional row needs to be preallocated. The length of each row is defined by the number of timestamps.

The robustness computation of the Always ( $\Box$ ), Eventually ( $\Diamond$ ), And ( $\wedge$ ), Or ( $\vee$ ), Not ( $\neg$ ) and Until ( $U$ ) operators is done in parallel in the C backend. The main program and interface of TLTK is implemented in Python. A wrapping function is utilized to ensure the execution of the C code from the Python interface. The algorithm and implementation details are presented in the extended technical report [7].



**Fig. 3.** A sample memory flow for the formula  $\phi = \Diamond(\neg r_1 \wedge r_2)$ . The initial memory allocation is shown in table (i) on the upper left corner. There are three rows allocated in total: two for the system output signals and one for the timestamps. In the next step, in table (ii), the memory that is allocated for the robustness calculations for  $r_1$  is utilized. Similarly, in the next two steps, in tables (iii and iv), there is a memory overwrite for the formula  $\neg r_1$  and robustness calculations for  $r_2$ . Next, in table (v), a memory overwrite occurs for the formula  $\neg r_1 \wedge r_2$  and a memory deallocation for the  $r_2$  distance row. Finally, in table (vi), there is a new memory allocation for the formula  $\Diamond(\neg r_1 \wedge r_2)$ , and memory deallocation for  $\neg r_1 \wedge r_2$ . The number of columns is divided by the number of available processor cores into groups. Each of these groups is processed in parallel.

## 4 Evaluation and Experimental Results

In the following, we evaluate the robustness computation times on various trace sizes and temporal logic specifications. We compare the outcome to the well-known tools DP-TALIRO [31] and BREACH [9]. In more detail, we compare TLTK with BREACH and DP-TALIRO separately so that we may focus in the areas where they have the strongest performance. The comparison with BREACH is focused on specifications with single-dimensional predicates, where the robustness of the predicates is computed through a subtraction of samples in a trajectory to the bound of the unsafe set. The comparison with DP-TALIRO is focused on specifications with multi-dimensional predicates, where the robustness computation for the predicates requires a solution to a quadratic program. Note that BREACH does not directly support robustness computations over multi-dimensional signal predicates. We highlight that in the following results, the TLTK robustness algorithm utilizes parallel processing, while the algorithms in DP-TALIRO and BREACH do not. We have verified the correctness of the robustness computation algorithm for each experimental result and, for ease of exposition, we focus the presentation on the computation time for the tools. The experiments were conducted on a desktop computer with the following specifica-

tions: CPU i7-8700K CPU @ 3.70GHz, GPU GTX 1080Ti, 32GiB DDR4 RAM and Ubuntu 18.04.3 LTS.

**Comparison with Breach (version 1.7)**<sup>6</sup>. We present the experimental comparison of TLTK with BREACH in Table 1. We compare the robustness computation of STL formulas with trace lengths ranging from  $2^{10}$  to  $2^{29}$ . The traces are checked against three STL specifications. A performance improvement of at least one order of magnitude can be observed.

$2^x$	$\varphi_{b1}$		$\varphi_{b2}$		$\varphi_{b3}$	
	TLTK	BREACH	TLTK	BREACH	TLTK	BREACH
10	0.00007	0.00350	0.00009	0.03343	0.00010	0.03588
14	0.00005	0.00935	0.00007	0.00480	0.00022	0.01243
18	0.00057	0.02322	0.00083	0.02862	0.00223	0.10138
22	0.00683	0.30040	0.01305	0.47182	0.03549	1.69170
26	0.10719	5.30410	0.21444	8.58840	0.56538	30.24000
27	0.21375	10.67300	0.42656	17.31700	1.12633	60.25600
28	0.42930	186.82000	0.85081	107.89000	2.24889	×
29	0.85353	×	1.69901	×	4.49088	×

Specification	Predicates
$\phi_{b1} = \neg(\Diamond s_1)$	$s_1 : speed(t) > 160$ $r_1 : rpm(t) < 4500$
$\phi_{b2} = \neg(\Diamond_{[0,1000]} s_1 \wedge \Box_{[100,300]} r_1)$	
$\phi_{b3} = \neg(\Diamond_{[0,1000]} s_1 \wedge \Box_{[0,200]} (r_1 \wedge \Box(\Diamond(s_1 \wedge (s_1 \mathcal{U} r_1))))$	

**Table 1.** Comparison of computation times in seconds for TLTK and BREACH with various specifications and trajectory lengths. × indicates out of memory error instances.

**Comparison with DP-TaLiRo (version 1.6)**<sup>7</sup>. The experimental comparison of TLTK with DP-TALiRo is presented in Table 2. The comparison is conducted using formulas that are defined for several benchmark problems. Specifically, requirements for the aircraftODE [27] ( $\phi_{s1}$ ), Navigation [27] ( $\phi_{s2}$ ), Heat Transfer [14] ( $\phi_{s6}$ ) and Nonlinear [1] ( $\phi_{s3-s5}$ ) Systems. The specifications are defined in Table 3. A performance improvement of at least two orders of magnitude can be observed.

## 5 Related Works

TLTK was inspired by the Matlab toolboxes S-TALiRo [3,11] and BREACH [9]. All three tools provide an automated test-case generation process for finding system behaviors that falsify temporal logic specifications. In addition to

<sup>6</sup> BREACH 1.7 downloaded on 01.16.2020 from <https://github.com/decyphir/breach>

<sup>7</sup> DP-TALiRo is part of S-TALiRo toolbox version 1.6. The tool was downloaded on 01.16.2020 from [https://app.assembla.com/spaces/s-taliro\\_public/](https://app.assembla.com/spaces/s-taliro_public/)



	$\varphi_{s1}$		$\varphi_{s2}$		$\varphi_{s3}$	
$2^x$	TLTK	DP-TALiRo	TLTK	DP-TALiRo	TLTK	DP-TALiRo
10	0.0023	1.5819	0.0028	1.5995	0.0018	1.8497
12	0.0087	6.3102	0.0106	6.3334	0.0081	7.0429
14	0.0295	25.1800	0.0361	25.3340	0.0252	28.1650
16	0.1118	100.6800	0.1375	101.3300	0.1002	112.6400
18	0.4429	403.1900	0.5334	405.1800	0.4013	450.4800
20	1.7296	1610.3000	2.1250	1621.0000	1.6054	1802.5000
21	3.5078	3222.3000	4.2977	3240.0000	3.2694	3604.3000
22	7.0906	×	8.5688	×	6.4353	×
23	14.0333	×	17.1810	×	13.0045	×
24	28.0057	×	34.2625	×	26.3092	×

	$\varphi_{s4}$		$\varphi_{s5}$		$\varphi_{s6}$	
$2^x$	TLTK	DP-TALiRo	TLTK	DP-TALiRo	TLTK	DP-TALiRo
10	0.0018	4.3029	0.0022	6.8373	0.0072	5.6698
12	0.0081	17.2180	0.0082	27.4090	0.0211	3.1871
14	0.0257	68.8750	0.0256	109.5900	0.0843	12.7620
16	0.1012	275.4100	0.1021	438.5200	0.3409	51.0760
18	0.4030	1102.0000	0.4097	1753.2000	1.3448	204.0600
20	1.6301	4402.8000	1.6199	7014.8000	5.4648	816.7000
21	3.2396	8805.1000	3.2225	14023.0000	10.9827	1632.2000
22	6.4321	×	6.4926	×	21.8416	×
23	12.8838	×	13.1343	×	43.2679	×
24	25.6910	×	26.0209	×	87.0204	×

**Table 2.** Comparison of computation times in seconds for TLTK and DP-TALiRo with various trajectory sizes. Specifications  $\varphi_{s1}$  through  $\varphi_{s6}$  and the predicate definitions can be found in Table 3. Symbol  $\times$  indicates out of memory instances.

falsification, these tools provide methods for requirement mining, conformance testing and real-time monitoring. They provide various optimization algorithms for black-box and grey-box testing. A different approach to falsification is utilized in the tool FALSTAR [32]. In FALSTAR, the falsifying system behavior is generated by constructing the input signal incrementally in time. This is particularly useful for reactive specifications. In another tool, **falsify** [2], the program solves the falsification problem through reinforcement learning. The tool attempts to find a falsification by observing the output signal and modifying the input signal during system simulation. A trajectory splicing/multiple shooting approach is utilized in the tool S3CAM [33], which explores the state-space of CPS and splices trace segments to find a path from one state to another. This approach was later extended to incorporate symbolic reachability techniques in the tool XSPEED [5]. To enable monitoring of temporal logic specifications in robotic sys-

MTL Specifications and Predicates	
$\phi_{s1} = \neg(\Box_{[5,150]}r_3 \wedge \Diamond_{[300,400]}r_4)$ $r_3 : A_{s1} * x \leq [250 \ -240]^T,$ $r_4 : A_{s1} * x \leq [240 \ -230]^T$	$A_{s1} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}$
$\phi_{s2} = (\neg p_{11})\mathcal{U}p_{12}$ where $p_{11} : A_{s2} * x \leq [3.8 \ -3.2 \ 0.8 \ -0.2]^T,$ $p_{12} : A_{s2} * x \leq [3.8 \ -3.2 \ 1.8 \ -1.2]^T$	$A_{s2} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix}$
$\phi_{s3} = \Box(r_7 \wedge \Diamond_{[0,100]}r_8)$ $\phi_{s4} = \neg(\Diamond r_7 \wedge \Box(r_8 \wedge \Box(\Diamond(r_7 \wedge (r_7\mathcal{U}r_8))))))$ $\phi_{s5} = \neg(\Diamond r_7 \wedge \Box(r_8 \wedge \Box(\Diamond(r_7 \wedge (r_7\mathcal{U}r_8)))) \wedge \Diamond(\Box(r_7 \vee (r_8\mathcal{U}r_7))))$ $r_7 : A_{s345} * x \leq [1.6 \ -1.4 \ 1.1 \ -0.9]^T,$ $r_8 : A_{s345} * x \leq [1.5 \ -1.2 \ 1.0 \ -1.0]^T$	$A_{s345} = \begin{bmatrix} -1 & 0 \\ 1 & 0 \\ 0 & -1 \\ 0 & 1 \end{bmatrix}$
$\phi_{s6} = \Box p$ $p : A_{s6} * x \leq [14.5 \ 14.5 \ 13.5 \ 14 \ 13 \ 14 \ 14 \ 13 \ 13.5 \ 14]^T$	$A_{s6} = I(10)$

**Table 3.** Specifications and predicates for the Signal Temporal Logic specifications utilized for the comparison between TLTK and S-TALiRo.

tems, in [26], the authors present RTAMT, which offers integration with ROS and supports monitoring of past and future time specifications.

## 6 Conclusion and Future Work

We have presented TLTK, a tool for falsification and parallel robustness computation of STL specifications. The modular architecture of the tool enables integration with any stochastic optimization algorithm or system simulator available in Python. The experimental results demonstrate that the multi-threaded CPU/GPU robustness engine shows a runtime improvement of at least one order of magnitude in comparison to BREACH and DP-TALiRo.

The robustness computation engine may be improved through syntactic analysis of the specifications to remove potentially redundant subformulas [8], or prioritizing results that imply results for other subformulas. In addition, as part of future work, the GPU algorithm may be improved further. In the current implementation, GPU computations are called for each predicate and temporal operator in the formula. This process causes an overhead when transferring the system trace to the GPU memory for each call. In addition, we plan to add requirement mining functionality as well as integration with ROS. Among our goals is to use the tool for planning and control of robotic systems.

## References

1. Abbas, H., Fainekos, G.E., Sankaranarayanan, S., Ivancic, F., Gupta, A.: Probabilistic temporal logic falsification of cyber-physical systems. *ACM Transactions on Embedded Computing Systems* **12**(s2) (May 2013)
2. Akazaki, T., Liu, S., Yamagata, Y., Duan, Y., Hao, J.: Falsification of cyber-physical systems using deep reinforcement learning. In: *Formal Methods. Lecture Notes in Computer Science*, vol. 10951, pp. 456–465. Springer (2018)
3. Annapureddy, Y.S.R., Liu, C., Fainekos, G.E., Sankaranarayanan, S.: S-taliro: A tool for temporal logic falsification for hybrid systems. In: *Tools and algorithms for the construction and analysis of systems. LNCS*, vol. 6605, pp. 254–257 (2011)
4. Bartocci, E., Deshmukh, J., Donzé, A., Fainekos, G., Maler, O., Ničković, D., Sankaranarayanan, S.: Specification-based monitoring of cyber-physical systems: a survey on theory, tools and applications. In: *Lectures on Runtime Verification*, pp. 135–175. Springer (2018)
5. Bogomolov, S., Frehse, G., Gurung, A., Li, D., Martius, G., Ray, R.: Falsification of hybrid systems using symbolic reachability and trajectory splicing. In: *Proceedings of the 22nd ACM International Conference on Hybrid Systems: Computation and Control*. pp. 1–10 (2019)
6. Cameron, F., Fainekos, G., Maahs, D.M., Sankaranarayanan, S.: Towards a verified artificial pancreas: Challenges and solutions for runtime verification. In: *Runtime Verification. LNCS*, vol. 9333, pp. 3–17. Springer (2015)
7. Cralley, J., Spantidi, O., Hoxha, B., Fainekos, G.: Tltk: Toolbox for parallel robustness computation of temporal logic specifications extended version (2020), online at: [http://bhoxha.com/papers/TLTK\\_ExtendedReport.pdf](http://bhoxha.com/papers/TLTK_ExtendedReport.pdf)
8. Dokhanchi, A., Hoxha, B., Fainekos, G.: Formal requirement debugging for testing and verification of cyber-physical systems. *ACM Transactions on Embedded Computing Systems (TECS)* **17**(2), 34 (2018)
9. Donze, A.: Breach, a toolbox for verification and parameter synthesis of hybrid systems. In: *Computer Aided Verification, LNCS*, vol. 6174, pp. 167–170 (2010)
10. Donze, A., Ferre, T., Maler, O.: Efficient robust monitoring for stl. In: *Proceedings of the 25th International Conference on Computer Aided Verification*. pp. 264–279. CAV, Springer-Verlag, Berlin, Heidelberg (2013)
11. Fainekos, G., Hoxha, B., Sankaranarayanan, S.: Robustness of specifications and its applications to falsification, parameter mining, and runtime monitoring with s-taliro. In: *International Conference on Runtime Verification*. pp. 27–47. Springer (2019)
12. Fainekos, G.E., Girard, A., Kress-Gazit, H., Pappas, G.J.: Temporal logic motion planning for dynamic robots. *Automatica* **45**(2), 343–352 (2009)
13. Fainekos, G.E., Pappas, G.J.: Robustness of temporal logic specifications for continuous-time signals. *Theoretical Computer Science* **410**(42), 4262–4291 (2009)
14. Fehnker, A., Ivancic, F.: Benchmarks for hybrid systems verification. In: *Hybrid Systems: Computation and Control. LNCS*, vol. 2993, pp. 326–341. Springer (2004)
15. Haghighi, I., Mehdipour, N., Bartocci, E., Belta, C.: Control from signal temporal logic specifications with smooth cumulative quantitative semantics. *arXiv preprint arXiv:1904.11611* (2019)
16. Hekmatnejad, M., Hoxha, B., Fainekos, G.: Search-based test-case generation by monitoring responsibility safety rules. In: *2020 IEEE Intelligent Transportation Systems Conference (ITSC)*. IEEE (2020)

17. Hoxha, B., Abbas, H., Fainekos, G.: Using s-taliro on industrial size automotive models. *Proc. of Applied Verification for Continuous and Hybrid Systems* (2014)
18. Hoxha, B., Abbas, H., Fainekos, G.: Benchmarks for temporal logic requirements for automotive systems. In: *Workshop on Applied Verification for Continuous and Hybrid Systems* (2014)
19. Hoxha, B., Dokhanchi, A., Fainekos, G.: Mining parametric temporal logic properties in model based design for cyber-physical systems. *International Journal on Software Tools for Technology Transfer* **20**, 79–93 (2018)
20. Jin, X., Deshmukh, J.V., Kapinski, J., Ueda, K., Butts, K.: Powertrain control verification benchmark. In: *To appear in Proc. of Hybrid Systems: Computation and Control* (2014)
21. Jin, X., Donzé, A., Deshmukh, J.V., Seshia, S.A.: Mining requirements from closed-loop control models. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* **34**(11), 1704–1717 (2015)
22. Kapinski, J., Deshmukh, J.V., Jin, X., Ito, H., Butts, K.: Simulation-based approaches for verification of embedded control systems: An overview of traditional and advanced modeling, testing, and verification techniques. *IEEE Control Systems* **36**(6), 45–64 (2016)
23. Leung, K., Aréchiga, N., Pavone, M.: Backpropagation for parametric stl. In: *2019 IEEE Intelligent Vehicles Symposium (IV)*. pp. 185–192. IEEE (2019)
24. Nelder, J.A., Mead, R.: A simplex method for function minimization. *The computer journal* **7**(4), 308–313 (1965)
25. Nghiem, T., Sankaranarayanan, S., Fainekos, G.E., Ivancic, F., Gupta, A., Pappas, G.J.: Monte-carlo techniques for falsification of temporal properties of non-linear hybrid systems. In: *Proceedings of the 13th ACM International Conference on Hybrid Systems: Computation and Control*. pp. 211–220. ACM Press (2010)
26. Nickovic, D., Yamaguchi, T.: Rtamt: Online robustness monitors from stl. *arXiv preprint arXiv:2005.11827* (2020)
27. Sankaranarayanan, S., Fainekos, G.: Falsification of temporal properties of hybrid systems using the cross-entropy method. In: *ACM International Conference on Hybrid Systems: Computation and Control* (2012)
28. Tuncali, C.E., Fainekos, G., Ito, H., Kapinski, J.: Simulation-based adversarial test generation for autonomous vehicles with machine learning components. In: *IEEE Intelligent Vehicles Symposium (IV)* (2018)
29. Tuncali, C.E., Hoxha, B., Ding, G., Fainekos, G., Sankaranarayanan, S.: Experience report: Application of falsification methods on the uxas system. In: *NASA Formal Methods Symposium*. pp. 452–459. Springer (2018)
30. Xiang, Y., Sun, D., Fan, W., Gong, X.: Generalized simulated annealing algorithm and its application to the thomson model. *Physics Letters A* **233**(3), 216–220 (1997)
31. Yang, H.: Dynamic Programming algorithm for Computing Temporal Logic Robustness. Master’s thesis, Arizona State University (2013)
32. Zhang, Z., Ernst, G., Sedwards, S., Arcaini, P., Hasuo, I.: Two-layered falsification of hybrid systems guided by monte carlo tree search. *IEEE Trans. on CAD of Integrated Circuits and Systems* **37**(11), 2894–2905 (2018)
33. Zutshi, A., Deshmukh, J.V., Sankaranarayanan, S., Kapinski, J.: Multiple shooting, cegar-based falsification for hybrid systems. In: *Proceedings of the 14th International Conference on Embedded Software*. pp. 1–10 (2014)