ARCH-COMP22 Category Report

EPiC Series in Computing

Volume 90, 2022, Pages 86–112

Proceedings of 9th International Workshop on Applied Verification of Continuous and Hybrid Systems (ARCH22)

ARCH-COMP22 Category Report: Continuous and Hybrid Systems with Nonlinear Dynamics

Luca Geretti[1], Julien Alexandre dit Sandretto[2], Matthias Althoff[3], Luis Benet[4], Pieter Collins[5], Parasara Sridhar Duggirala[6], Marcelo Forets[7], Edward Kim[6], Stefan Mitsch[8], Christian Schilling[9], and Mark Wetzlinger[3]

[1] Department of Computer Science, University of Verona, Verona, Italy
luca.geretti@univr.it
[2] ENSTA Paris, Institut Polytechnique de Paris, Palaiseau, France
julien.alexandre-dit-sandretto@ensta-paris.fr
[3] Technische Universität München, Munich, Germany
althoff@in.tum.de,m.wetzlinger@tum.de
[4] Instituto de Ciencias Físicas, Universidad Nacional Autónoma de México (UNAM), México
benet@icf.unam.mx
[5] Department of Data Science and Knowledge Engineering, Maastricht University, Maastricht, The Netherlands
pieter.collins@maastrichtuniversity.nl
[6] Department of Computer Science, University of North Carolina at Chapel Hill, US
psd@cs.unc.edu,ehkim@cs.unc.edu
[7] Universidad de la República, Montevideo, Uruguay
mforets@gmail.com
[8] Computer Science Department, Carnegie Mellon University, Pittsburgh, PA, USA
smitsch@cs.cmu.edu
[9] Aalborg University, Aalborg, Denmark
christianms@cs.aau.dk

Abstract

We present the results of a friendly competition for formal verification of continuous and hybrid systems with nonlinear continuous dynamics. The friendly competition took place as part of the workshop Applied Verification for Continuous and Hybrid Systems (ARCH) in 2022. This year, 6 tools Ariadne, CORA, DynIbex, JuliaReach, Kaa and KeYmaera X (in alphabetic order) participated. These tools are applied to solve reachability analysis problems on six benchmark problems, two of them featuring hybrid dynamics. We do not rank the tools based on the results, but show the current status and discover the potential advantages of different tools.

1 Introduction

Disclaimer The presented report of the ARCH friendly competition for *continuous and hybrid systems with nonlinear dynamics* aims at providing a landscape of the current capabilities of verification tools. We would like to stress that each tool has unique strengths—though not all of their features can be highlighted within a single report. To reach a consensus in what benchmarks are used, some compromises had to be made so that some tools may benefit more from the presented choice than others. The obtained results have been verified by an independent repeatability evaluation. To establish further trustworthiness of the results, the code with which the results have been obtained is publicly available as Docker [17] containers at gitlab.com/goranf/ARCH-COMP.

In this report, we summarize the results of the sixth ARCH friendly competition on the reachability analysis of continuous and hybrid systems with nonlinear dynamics. Given a system defined by a nonlinear Ordinary differential equation (ODE) $\dot{\vec{x}} = f(\vec{x}, t)$ along with an initial condition $\vec{x} \in X_0$, we apply the participating tools to prove properties of the state reachable set in a bounded time horizon. The techniques for solving such a problem are usually very sensitive to not only the nonlinearity of the dynamics but also the size of the initial set. This is also one of the main reasons why most of the tools require quite a lot of computational parameters.

In this report, 6 tools, namely Ariadne, CORA, DynIbex, JuliaReach, Kaa and KeYmaera X participated in solving problems defined on three continuous and two hybrid benchmarks. This year the KeYmaera X tool joined the competition. The continuous benchmarks are the Traffic scenario, the Robertson chemical reaction system, the Coupled Van der Pol oscillator and the Laub-Loomis model of enzymatic activities. The hybrid benchmarks model a Lotka-Volterra predator-prey system with a Tangential Crossing, and a Space Rendezvous system.

The benchmarks were selected based on discussions between the tool authors, with a preference on keeping a significant set of the benchmarks from the previous year. It is apparent that they come from very different domains and aim at identifying issues specific to nonlinear dynamics, possibly with the addition of hybrid behavior.

2 Participating Tools

Ariadne. (Luca Geretti, Pieter Collins) *Ariadne* [23, 16] is a library based on *Computable Analysis* [48] that uses a rigorous numerical approach to all its algebraic, geometric and logical operations. In particular, it performs numerical rounding control of all external and internal operations, in order to enforce conservative interpretation of input specification and guarantee formal correctness of the computed output. It focuses on nonlinear systems, both continuous and hybrid, supporting differential and algebraic relations, with a focus on compositionality [20]. It has been mainly applied to the verification of robotic surgery tasks [21]. The library is written in modern C++ with an optional Python interface. The official site for Ariadne is `https://www.ariadne-cps.org`.

CORA. (Matthias Althoff, Mark Wetzlinger) The tool *COntinuous Reachability Analyzer* (CORA) [6, 7] realizes techniques for reachability analysis with a special focus on developing scalable solutions for verifying hybrid systems with nonlinear continuous dynamics and/or nonlinear differential-algebraic equations. A further focus is on considering uncertain parameters and system inputs. Due to the modular design of CORA, much functionality can be used for other purposes that require resource-efficient representations of multi-dimensional sets and operations on them. CORA is implemented as an object-oriented MATLAB code. The modular design of CORA makes it possible to use the capabilities of the various set representations for other purposes besides reachability analysis. While CORA uses verified algorithms, it does not consider rounding errors since the main focus of the toolbox is the fast prototyping of new reachability algorithms and concepts, and for this purpose the effect of rounding errors is usually negligible. CORA is available at cora.in.tum.de.

DynIbex. (Julien Alexandre dit Sandretto) A library merging interval constraint satisfaction problem algorithms and guaranteed numerical integration methods based on Runge-Kutta numerical schemes implemented with affine arithmetic. This library is able to solve ordinary differential equations [2] and algebraic differential equations of index 1 [3], combined with numerical constraints on state variables and reachable tubes. It produces **sound results** taking into account round-off errors in floating-point computations and truncation errors generated by numerical integration methods [39]. Moreover, constraint satisfaction problem algorithms offer a convenient approach to check properties on reachable tubes as explained in [4]. This library implements in a very generic way validated numerical integration methods based on Runge-Kutta methods without many optimizations. Indeed, the computation of the local truncation error, for each method, depends only on the coefficients of Runge-Kutta methods and their order. DynIbex is freely available at `http://perso.ensta-paristech.fr/~chapoutot/dynibex/`. Figures have been produced with VIBes library [26] which is available at `http://enstabretagnerobotics.github.io/VIBES/`. Computations are performed on a Lenovo laptop with i5 processor, and computation times gather all the process from compilation to figure producing.

JuliaReach. (Luis Benet, Marcelo Forets, Christian Schilling) JuliaReach [18] is an open-source software suite for reachability computations of dynamical systems, written in the Julia language and available at `http://github.com/JuliaReach`. Linear, nonlinear, and hybrid problems are modeled and solved using the library ReachabilityAnalysis.jl, which can be used interactively, for example in Jupyter notebooks. Our implementation of the Taylor-model based solvers, (`TMJets20`, `TMJets21a`, and `TMJets21b`), which are implemented in TaylorModels.jl [15], integrates the packages TaylorSeries.jl [12, 13] and TaylorIntegration.jl [40], and the IntervalArithmetic.jl [14] package for interval methods. The algorithms applied in this report first compute a non-validated integration using a Taylor model of order n_T. The coefficients of that series are polynomials of order n_Q in the variables that denote small deviations of the initial conditions. We obtain a time step from the last two coefficients of this time series. In order to validate the integration step, we compute a second integration using intervals as coefficients of the polynomials in time, and we obtain a bound for the integration using a Lagrange-like remainder. The remainder is used to check the contraction of a Picard iteration. If the combination of the time step and the remainder do not satisfy the contraction, we iteratively enlarge the remainder or possibly shrink the time step. Finally, we evaluate the initial Taylor series with

the valid remainder at the time step for which the contraction has been proved, which is also evaluated in the initial set to yield an over-approximation. The approach is (numerically) sound due to rigorous interval bounds in the Taylor approximation. Discrete transitions for hybrid systems and Taylor-model approximations are handled using the set library LazySets.jl [27].

Kaa. (Parasara Sridhar Duggirala, Edward Kim) Kaa [32] is a Python rewrite of Sapo [24], a tool written to compute and plot the reachable sets of discrete polynomial non-linear dynamical systems using parallelotope bundles. The representation of parallelotope bundles for reachability was proposed in [25]. We extend these techniques to exploit dynamic templates strategies, i.e schemes created to automatically generate template directions and parallelotopes. We use two techniques to generate such template directions. The first involves computing *local linear approximations* of the dynamics, and the second involves performing *Principle Component Analysis (PCA)*. Both techniques are performed using sample trajectory data. The sample trajectories are found by calculating *support points* over the parallelotope bundle and propagating them to the next step using the provided dynamics. Each parallelotope added to the bundle has an associated *lifespan*, indicating the number of time steps the parallelotope and its template directions exist in the bundle before being removed. We employ a Python wrapper over NASA's Kodiak[1] as the optimization library responsible for computing the upper and lower offsets for all utilized template directions at each step. The original Sapo program could only handle polynomial dynamics through Bernstein polynomials. However, Kodiak allows us to circumvent this restriction and extend our techniques to general non-linear dynamics. Sympy is used for symbolic manipulation and substitution while Numpy is used for general linear-algebraic computations. Detailed explanations of the underlying techniques used in Kaa can be found in a recent paper including the co-authors [31].

KeYmaera X. (Stefan Mitsch) KeYmaera X [29] is a theorem prover for the hybrid systems logic differential dynamic logic (dL). It implements the uniform substitution calculus of dL [41]. A comparison of the internal reasoning principles in the KeYmaera family of provers with a discussion of their relative benefits and drawbacks is in [38], and model structuring and proof management on top of uniform substitution is discussed in [36]. KeYmaera X supports systems with nondeterministic discrete jumps, nonlinear differential equations, nondeterministic inputs, and allows defining functions implicitly through their characterizing differential equations [30]. It provides invariant construction and proving techniques for differential equations [45, 42], and stability verification techniques for switched systems [46]. Unlike numerical hybrid systems reachability analysis tools, KeYmaera X also supports unbounded initial sets and unbounded time analysis. Proofs in KeYmaera X can be conducted interactively [37], steered with tactics [28], or attempted fully automatic.

3 Benchmarks

For the 2022 edition of the competition we introduced one new continuous benchmark: the *Traffic scenario* system. This is a continuous system with disturbances and piecewise-constant input that, for the first time, introduces time-varying dynamics into the mix. In addition, we modified the *van der Pol* and the *Space Rendezvous* system to make them slightly more difficult compared to last year. The *Robertson*, *Laub-Loomis* and *Lotka-Volterra* systems were not modified, in order to identify any improvements to the tools from 2021. Differently from previous years, no benchmark has been dropped entirely.

3.1 Traffic scenario benchmark (TRAF22)

The avoidance of collisions in traffic scenarios is of utmost interest in the development of motion planners for autonomous driving. Recently [34], a workflow for the automated generation of verification tasks has been proposed based on an extraction of traffic scenario benchmarks from the CommonRoad framework [8].

3.1.1 Model

The nonlinear continuous-time dynamics are represented by a kinematic single-track model [34, Eq. (1)]:

$$\begin{cases} \dot{\delta} = u_1 + w_1 \\ \dot{\psi} = \dfrac{v}{l_{wb}} \tan\delta \\ \dot{v} = u_2 + w_2 \\ \dot{s}_x = v\cos\psi \\ \dot{s}_y = v\sin\psi, \end{cases}$$

where the state vector $x \in \mathbb{R}^5$ consists of the steering angle δ, the vehicle heading ψ, the vehicle velocity v, and the positions s_x, s_y of the vehicle along the x-axis and y-axis. The control inputs u_1, u_2 represent the steering angle and acceleration, respectively. Additionally, model uncertainties and disturbances affecting the vehicle are modeled by the disturbances w_1, w_2. In order to follow a reference trajectory $x_{ref} \in \mathbb{R}^5$, we apply a feedback controller of the form [34, Eq. (2)]

$$u_{fb}(\widehat{x}) = u_{ref} + K(\widehat{x} - x_{ref})$$

with the time-varying reference input $u_{ref} \in \mathbb{R}^2$, the time-varying feedback matrix $K \in \mathbb{R}^{2\times 5}$, and the measured state $\widehat{x} := x + v$ defined using the measurement error $v \in \mathbb{R}^5$. Thus, the ten-dimensional closed-loop system $f(x, u, w)$ is obtained by inserting the control law into the five-dimensional model:

$$\begin{cases} \dot{x} = f(x, u_{ref} + K(x + v - x_{ref}), w) \\ \dot{x}_{ref} = f(x_{ref}, u_{ref}, 0) \end{cases}$$

Table 1: Results of TRAF22 in terms of computation time and verification.

tool	**computation time in [s]**	**Verified?**
Ariadne	N/A	N/A
CORA	26	Yes
DynIbex	N/A	N/A
JuliaReach	N/A	N/A
Kaa	N/A	N/A
KeYmaera X	N/A	N/A

3.1.2 Analysis

The set for the measurement error $\mathcal{V} \subset \mathbb{R}^5$, the input set $\mathcal{U} \subset \mathbb{R}^2$, and the set of disturbances $\mathcal{W} \subset \mathbb{R}^2$ are respectively bounded by

$$\mathcal{V} = \begin{pmatrix} [-0.0004, 0.0004] \\ [-0.0004, 0.0004] \\ [-0.006, 0.006] \\ [-0.002, 0.002] \\ [-0.002, 0.002] \end{pmatrix} \qquad \mathcal{U} = \begin{pmatrix} [-0.7, 0.7] \\ [-11, 11] \end{pmatrix} \qquad \mathcal{W} = \begin{pmatrix} [-0.02, 0.02] \\ [-0.3, 0.3] \end{pmatrix}.$$

The initial state is uncertain within the set $x_0 \oplus \mathcal{V} \times x_0$. The inputs u_1, u_2 and the disturbances w_1, w_2 can change arbitrarily over time within their respective sets.

In this case, we analyze the scenario with the identifier *BEL_Putte-4_2_T-1*: The time horizon is determined by the length of the piecewise-constant control values, i.e., the reference trajectory x_{ref}, reference input u_{ref}, and feedback matrix K. All of these are provided by a .csv-file in a format as detailed in [34, Sec. 5].

The following two specifications have to be satisfied:

- **Input constraints**: The controller input $u_{fb} \in \mathbb{R}^2$ should be contained within the input set $\mathcal{U}$ at all times. The set of control inputs is computed according to [34, Eq. (5)].

- **Collision avoidance**: The car should not collide with static or dynamic obstacles as well as the road boundaries. Therefore, one requires to compute the car's occupancy set according to [34, Eq. (4)]. After rewriting the occupancy set as a .csv-file using the format in [34, Fig. 4], the collision check is performed fully automatically by calling a provided Python script as detailed in [34, Sec. 5].

3.1.3 Evaluation

There are two metrics to evaluate the performance of each tool: First, we measure the computation time only comprising the time spent during the reachable set computation, exempt the time step in the pre- and post-processing steps. Second, we explicitly tabulate the results of the verification since a collision could occur at any time and therefore might not be captured in the figures below.

3.1.4 Results

The results from this benchmark are shown in Table 1. Except for CORA, all the tools currently do not support the format required by the benchmark. Hence a proper comparison is not possible and the benchmark will be re-proposed as-is next year.

Settings for Ariadne. Ariadne is currently able to express disturbances within purely continuous dynamics, while the piecewise-constant input requires extension to the hybrid space. We plan on supporting hybrid systems for the next year.

Settings for CORA. We used the conservative linearization approach [10] with a time step size of $\Delta t = 0.005$, resulting in 20 steps per piecewise-constant input. Despite the relatively high system dimension, a zonotope order of 20 was sufficient for successful verification.

Settings for DynIbex. DynIbex does not participate to this benchmark this year.

Settings for JuliaReach. JuliaReach does not currently support this model format, but support is planned for next year's edition.

Settings for Kaa. Kaa does not feature any support to model uncertainty or disturbances as specified by this model.

Settings for KeYmaera X. KeYmaera X does not currently support reading streams of recorded control inputs and outputting occupancy sets as required in the collision avoidance simulation step of this benchmark. In future editions, we plan to formalize streams of control inputs fully symbolically to characterize the safety-relevant properties of such streams and conduct proofs for any control input stream satisfying these properties.

3.2 Robertson chemical reaction benchmark (ROBE21)

3.2.1 Model

As proposed by Robertson [43], this chemical reaction system models the kinetics of an autocatalytic reaction.

$$\begin{cases} \dot{x} = -\alpha x + \beta yz \\ \dot{y} = \alpha x - \beta yz - \gamma y^2 \\ \dot{z} = \gamma y^2 \end{cases}$$

where x, y and z are the (positive) concentrations of the species, with the assumption that $x+y+z = 1$. Here α is a small constant, while β and γ take on large values. In this benchmark we fix $\alpha = 0.4$ and analyze the system under three different pairs of values for β and γ:

1. $\beta = 10^2$, $\gamma = 10^3$
2. $\beta = 10^3$, $\gamma = 10^5$
3. $\beta = 10^3$, $\gamma = 10^7$

The initial condition is always $x(0) = 1$, $y(0) = 0$ and $z(0) = 0$.

3.2.2 Analysis

We are interested in computing the reachable tube until $t = 40$, to see how the integration scheme holds under the stiff behavior. No verification objective is enforced.

3.2.3 Evaluation

For each of the three setups, the following three measures are required:

1. the execution time for evolution;
2. the number of integration steps taken;
3. the width of the sum of the concentrations $s = x + y + z$ at the final time.

Additionally, a figure with s (in the $[0.999, 1.001]$ range) w.r.t. time overlaid for the three setups should be shown.

3.2.4 Results

Except for Kaa, tools were able to get to completion. However, very different results were obtained. In the case of Ariadne and JuliaReach, the width started small and increased monotonically, while for DynIbex and CORA the width started decreasing from a given value, to possibly increase further in the case of DynIbex. It is also interesting to analyze the number of integration steps taken, which turned out to be sensibly lower for JuliaReach and especially CORA. While JuliaReach obtained the best width for the stiffer cases, this came at the expense of a significantly higher computation time. Perhaps for the next year some verification constraints should be enforced, in order to provide a better baseline for comparison between the tools.

Settings for Ariadne. A GradedTaylorSeriesIntegrator is used, with a maximum error per integration step of 10^{-9}. A maximum step size of 0.004 is imposed in all three setups, though the actual value dynamically identified along evolution for (2) and (3) is sensibly lower.

Settings for CORA. In all cases, we used the conservative linearization approach [10] with maximum zonotope orders of 30, 30, and 200, respectively. In order to accelerate the computation, we increased the time step size during the computation, at the loss of some amount of tightness of the reachable sets. The used time step sizes are (1) $t \in [0, 5] : 0.0025, t \in [5, 40] : 0.025$, (2) $t \in [0, 5] : 0.002, t \in [5, 40] : 0.005$, and (3) $t \in [0, 2] : 0.0002, t \in [2, 40] : 0.001$.

Settings for DynIbex. The Runge-Kutta method selected is implicit Lobatto at fourth order (called LC3 in DynIbex) for the three setups. The absolute precision is respectively 10^{-11}, 10^{-12} and 10^{-12}. The other parameters are set by default.

Settings for JuliaReach. In all cases we use $n_Q = 1$, an initial adaptive absolute tolerance 10^{-10} and the `TMJets21a` algorithm, adapting only the n_T parameter as follows: (1) $n_T = 5$, (2) $n_T = 7$ and (3) $n_T = 10$. The maximum number of integration steps is also adjusted, reflecting the results presented in Table 2. For the results displayed in Fig. 1, we evaluate s directly on the Taylor models produced by the integration.

Table 2: Results of ROBE21 in terms of computation time, number of steps and width of $s = x + y + z$.

	computation time in [s]		
tool	(1)	(2)	(3)
Ariadne	35	194	371
CORA	35	67	401
DynIbex	309	3426	5204
JuliaReach	69	1120	4544
Kaa	–	–	–
KeYmaera X[1]	0.2	0.2	0.2

[1] Single symbolic proof solves all 3 examples

	number of steps		
tool	(1)	(2)	(3)
Ariadne	10000	49849	123675
CORA	3400	9500	48000
DynIbex	8694	84460	123248
JuliaReach	3494	30147	71367
Kaa	–	–	–
KeYmaera X[1]	113	113	113

[1] Single symbolic proof solves all 3 examples

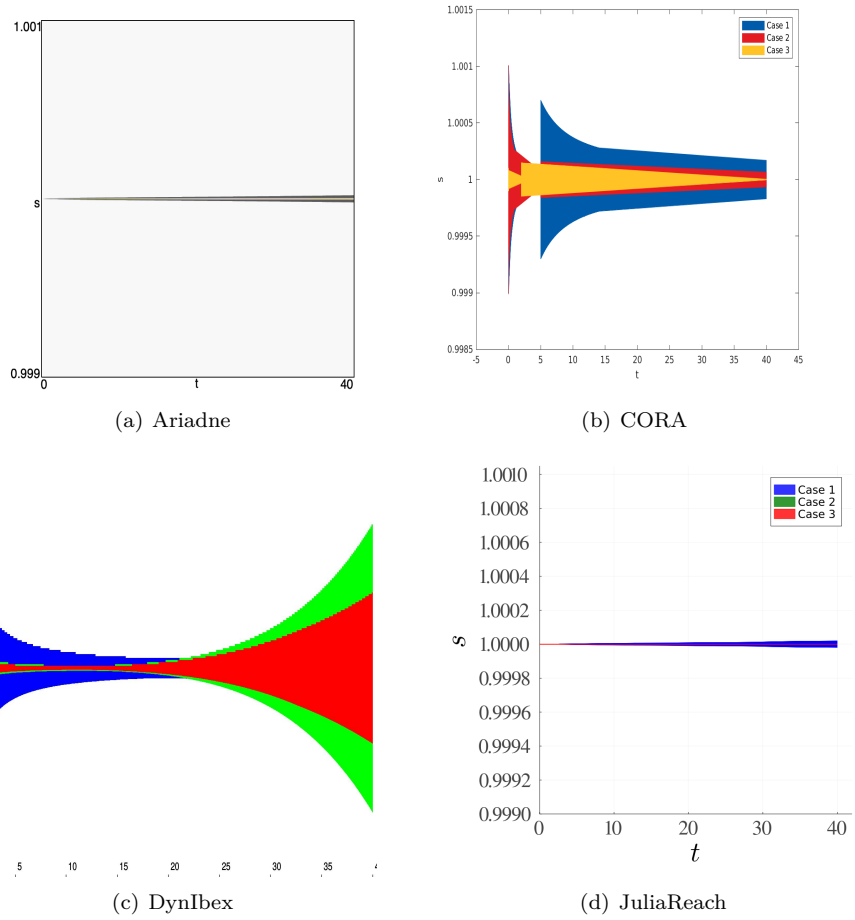
	width of $x + y + z$		
tool	(1)	(2)	(3)
Ariadne	1.0e-5	4.0e-5	8.0e-6
CORA	3.4e-4	1.3e-4	7.1e-6
DynIbex	7.9e-4	1.4e-3	7.6e-4
JuliaReach	3.8e-5	8.1e-8	1.2e-9
Kaa	–	–	–
KeYmaera X[1]	0	0	0

[1] Exact computation without overapproximation

Settings for Kaa. Similar to last year, the dynamics causes explosive growth within the beginning of the computation; this explosion generally occurs between $t \in [0, 2]$. Our techniques, both static and dynamic, seem to be unable to control this dramatic behavior, which causes Kodiak to eventually crash. We attempted this year to use Bernstein polynomials with no dynamic templates to study the behavior of the reachability algorithm during the first few steps. The coefficients of the polynomials we attempt to optimize over explode in value. This is most likely due to values of β, γ fixed in all three cases. In light of this, it is possible that optimizing over the polynomials after functional composition is unfit for this benchmark.

Settings for KeYmaera X. The KeYmaera X proof is fully parametric, without approximation, and shows stability of all possible population sums s for any (even negative) choice of a, b, and g, which includes the specific parametrizations (i) $b = 10^2, g = 10^3$, (ii) $b = 10^3, g = 10^5$, and (iii) $b = 10^3, g = 10^7$.

```
1  Problem
2    x+y+z=s
3    ->
4    [{ x' = -a*x + b*y*z,
5       y' = a*x - b*y*z - g*y^2,
6       z' = g*y^2
7     }
8    ](x+y+z=s)
9  End.
10
```



(a) Ariadne (b) CORA

(c) DynIbex (d) JuliaReach

Figure 1: Reachable set overapproximations of $s = x + y + z$ vs time for ROBE21 in the three setups.

```
11 Tactic "Scripted proof" unfold; dIClose(1) End.
12 Tactic "Automated proof" autoClose End.
```

3.3 Coupled van der Pol benchmark (CVDP22)

3.3.1 Model

The original van der Pol oscillator was introduced by the Dutch physicist Balthasar van der Pol. For this benchmark we consider two coupled oscillators, as described in [11]. The system can be defined by the following ODE with 5 variables:

$$\begin{cases} \dot{x}_1 &= y_1 \\ \dot{y}_1 &= \mu(1-x_1^2)y_1 + b(x_2 - x_1) - x_1 \\ \dot{x}_2 &= y_2 \\ \dot{y}_2 &= \mu(1-x_2^2)y_2 - b(x_2 - x_1) - x_2 \\ \dot{b} &= 0 \end{cases} \tag{1}$$

with $\mu = 1$. The system has a stable limit cycle that becomes increasingly sharper for higher values of μ.

3.3.2 Analysis

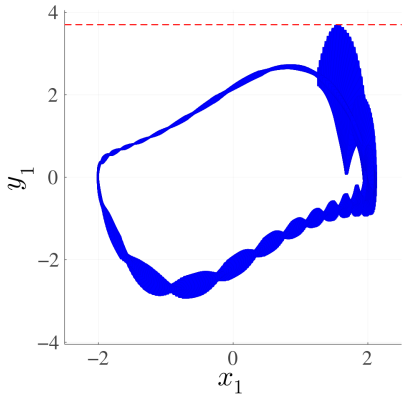
We set the initial condition $x_{1,2}(0) \in [1.25, 1.55]$, $y_{1,2}(0) \in [2.35, 2.45]$ and $b \in [60\,, 80]$. The unsafe set is given by $y_{1,2} \geq 3.7$ in a time horizon of $[0, 7]$.

3.3.3 Evaluation

The computation time required to evolve the system and verify safety is provided. If the system can not be verified successfully, no value is given.

3.3.4 Results

The computation results of the tools are given in Table 3. While DynIbex and KeYmaera X were not able to participate in this specific benchmark, Ariadne, Cora and Kaa encountered numerical problems that prevented completion in a reasonable time. Only JuliaReach was able to address the benchmark properly.



(a) JuliaReach

Settings for Ariadne. It was not possible to achieve completion is a reasonable time, due to the very high number of splittings theoretically required to guarantee numerical convergence.

Settings for CORA. Due to the strong nonlinearity introduced by the parameter b, CORA was not able to produce satisfactory results without splitting the initial set in too many subsets.

Table 3: Results of CVDP22 in terms of computation time.

tool	**computation time in [s]**
Ariadne	–
CORA	–
DynIbex	N/A
JuliaReach	3.6
Kaa	–
KeYmaera X	N/A

Settings for DynIbex. DynIbex does not participate to this benchmark this year.

Settings for JuliaReach. We use $n_Q = 1$, $n_T = 8$, and an adaptive absolute tolerance 10^{-10}.

Settings for Kaa. The step sizes for both cases were set to $\Delta = 0.1$. Once again, we encountered severe numerical instabilities, especially from local linear approximations arising from computing the inverse of matrices with high condition numbers. The extra b dimension greatly lengthens the computation time for the full set of Bernstein coefficients. Furthermore, as seen last year, both Kodiak and static parallelotopes result in segmentation fault. The dynamics and initial parameters seem to cause great variance in the optimization solutions produced by Kodiak. This results in the reachable set plot for early steps $t \in [0, 1]$ to appear extremely malformed and jagged. Hence, we believe that our current method of choosing reasonable templates is ill-suited for producing meaningful over-approximations for this benchmark.

Settings for KeYmaera X. The Coupled van der Pol benchmark was formalized for KeYmaera X but not yet proved. Below, we give the formal specification in KeYmaera X format:

```
1  Definitions Real m, b; End.
2  ProgramVariables Real x1, x2, y1, y2; End.
3  Problem
4      60 <= b & b <= 80                                              /* b in [60,80] */
5    & m = 1
6    & 1.25 <= x1 & x1 <= 1.55 & 1.25 <= x2 & x2 <= 1.55 /* x_{1,2}(0) in [1.25,1.55] */
7    & 2.35 <= y1 & y1 <= 2.45 & 2.35 <= y2 & y2 <= 2.45 /* y_{1,2}(0) in [2.35,2.45] */
8    & t = 0
9    ->
10   [{ x1' = y1,
11      y1' = m*(1-x1^2)*y1 + b*(x2-x1) - x1,
12      x2' = y2,
13      y2' = m*(1-x2^2)*y2 - b*(x2-x1) - x2,
14      t' = 1 & t <= 7                                          /* time horizon [0,7] */
15    }
16   ]!( y1>=3.7 & y2>=3.7)                                      /* not in unsafe set */
17 End.
```

In future editions, we plan to search and prove correct symbolic invariant conditions of the dynamics.

3.4 Laub-Loomis benchmark (LALO20)

3.4.1 Model

The Laub-Loomis model is presented in [35] for studying a class of enzymatic activities. The dynamics can be defined by the following ODE with 7 variables.

$$\begin{cases} \dot{x}_1 &= 1.4x_3 - 0.9x_1 \\ \dot{x}_2 &= 2.5x_5 - 1.5x_2 \\ \dot{x}_3 &= 0.6x_7 - 0.8x_2x_3 \\ \dot{x}_4 &= 2 - 1.3x_3x_4 \\ \dot{x}_5 &= 0.7x_1 - x_4x_5 \\ \dot{x}_6 &= 0.3x_1 - 3.1x_6 \\ \dot{x}_7 &= 1.8x_6 - 1.5x_2x_7 \end{cases}$$

The system is asymptotically stable and the equilibrium is the origin.

3.4.2 Analysis

The specification for the analysis is kept the same as last year, in order to better quantify any improvements to the participating tools.

The initial sets are defined according to the ones used in [47]. They are boxes centered at $x_1(0) = 1.2$, $x_2(0) = 1.05$, $x_3(0) = 1.5$, $x_4(0) = 2.4$, $x_5(0) = 1$, $x_6(0) = 0.1$, $x_7(0) = 0.45$. The range of the box in the ith dimension is defined by the interval $[x_i(0) - W, x_i(0) + W]$. The width W of the initial set is vital to the difficulty of the reachability analysis job. The larger the initial set the harder the reachability analysis.

We consider $W = 0.01$, $W = 0.05$, and $W = 0.1$. For $W = 0.01$ and $W = 0.05$ we consider the unsafe region defined by $x_4 \geq 4.5$, while for $W = 0.1$, the unsafe set is defined by $x_4 \geq 5$. The time horizon for all cases is $[0, 20]$.

3.4.3 Evaluation

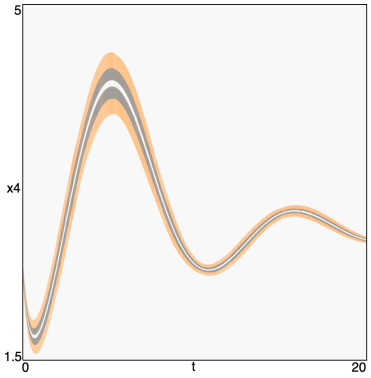
The final widths of x_4 along with the computation times are provided for all three cases. A figure is provided in the (t, x_4) axes, with $t \in [0, 20]$, $x_4 \in [1.5, 5]$, where the three plots are overlaid.

3.4.4 Results

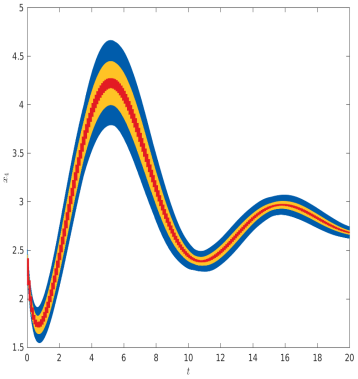
The computation results of the tools are given in Table 4. Results are essentially the same as last year's.

Settings for Ariadne. The maximum step size used is 0.2, with a TaylorPicardIntegrator with a maximum spacial error of 10^{-6} enforced for each step. Compared with last year, no splitting strategy for the initial set is necessary due to improvements in the integrator.

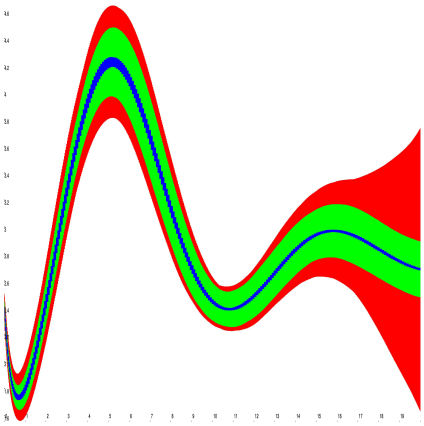
Settings for CORA. Depending on the size of the initial set, different algorithms in CORA are applied. For the smaller initial sets $W = 0.01$ and $W = 0.05$, the faster but less accurate conservative linearization algorithm presented in [10] is executed. For the larger initial set $W = 0.1$, the more accurate conservative polynomialization algorithm from [5] is applied. CORA uses a step size of 0.1 for $W = 0.01$, a step size of 0.025 for $W = 0.05$, and a step size of 0.02 for $W = 0.1$. For all sizes of initial sets, the maximum zonotope order is chosen as 200.



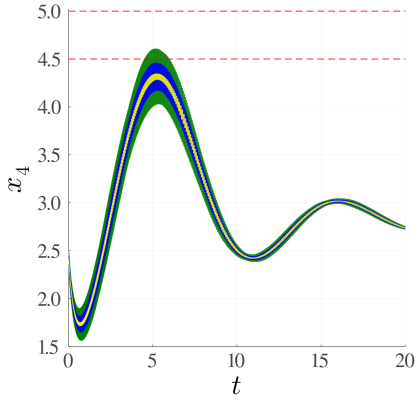
(b) Ariadne



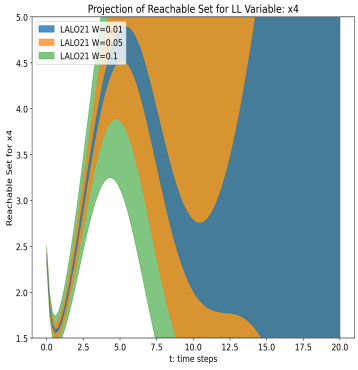
(c) CORA



(d) DynIbex



(e) JuliaReach



(f) Kaa

Figure 2: Reachable set overapproximations for LALO20 (overlayed plots for $W = 0.01$, $W = 0.05$, $W = 0.1$). $t \in [0, 20]$, $x_4 \in [1.5, 5]$.

Table 4: Results of LALO20 in terms of computation time and width of final enclosure.

	computation time in [s]		
tool	$W = 0.01$	$W = 0.05$	$W = 0.1$
Ariadne	6.8	20	47
CORA	2.2	9.5	47
DynIbex	10	27	1909
JuliaReach	4.0	5.8	5.8
Kaa	336	355	363
KeYmaera X	NA	NA	NA

	width of x_4 in final enclosure		
tool	$W = 0.01$	$W = 0.05$	$W = 0.1$
Ariadne	0.009	0.030	0.070
CORA	0.005	0.035	0.116
DynIbex	0.01	0.40	2.07
JuliaReach	0.0042	0.017	0.033
Kaa	23	29	46
KeYmaera X	NA	NA	NA

Settings for DynIbex. For $W = 0.01$ the maximum zonotope order is set to 50 and the reachability analysis is carried out with an (absolute and relative) error tolerance of 10^{-6} with an explicit Runge-Kutta method of order 3. For $W = 0.05$ the maximum zonotope order is set to 80 and the reachability analysis is carried out with an (absolute and relative) error tolerance of 10^{-7} with an explicit Runge-Kutta method of order 3. For $W = 0.01$ and $W = 0.05$ no splitting of the initial conditions is performed. For $W = 0.1$, the initial set is split 64 times. With parallelization, computation time is reduced to 249 seconds for this last experiment.

Settings for JuliaReach. We use an absolute tolerance of 10^{-11} for $W = 0.01$ and 10^{-12} for $W = 0.05$ and $W = 0.1$. In all cases, $n_Q = 1$ and $n_T = 7$.

Settings for Kaa. For each case, we employed a dynamic strategy of PCA templates with their lifespan set to 10 steps. Similar to the previous year, we faced difficulties with using local linear approximation. However,from the plots, PCA templates are ineffective in controlling the error as time progresses. Due to this wrapping error, we capped the upper and lower offsets for each template direction to lie in between the intervals $[-10, 10]$ in order to ensure completion of the computation. We used a step size of $\Delta = 0.2$ for all cases.

Settings for KeYmaera X. The Laub-Loomis benchmark was formalized for KeYmaera X but not yet proved. Below, we give the formal specification in KeYmaera X format:

```
1 Definitions
2   Real W = 0.1;
3   Bool box(Real x, Real c, Real w) <-> c-w <= x & x <= c+w;
4 End.
```

```
5
6   ProgramVariables
7     Real x1, x2, x3, x4, x5, x6, x7;  /* state space */
8     Real t;                           /* time */
9   End.
10
11  Problem
12      box(x1, 1.2, W)                 /* initial sets */
13    & box(x2, 1.05, W)
14    & box(x3, 1.5, W)
15    & box(x4, 2.4, W)
16    & box(x5, 1, W)
17    & box(x6, 0.1, W)
18    & box(x7, 0.45, W)
19    & t=0
20    ->
21    [{ x1' = 1.4*x3 - 0.9*x1,
22       x2' = 2.5*x5 - 1.5*x2,
23       x3' = 0.6*x7 - 0.8*x2*x3,
24       x4' = 2 - 1.3*x3*x4,
25       x5' = 0.7*x1 - x4*x5,
26       x6' = 0.3*x1 - 3.1*x6,
27       x7' = 1.8*x6 - 1.5*x2*x7,
28       t'  = 1 & t<=20              /* time horizon [0,20] */
29     }
30    ]!( x4>=5)                      /* not in unsafe set */
31  End.
```

In future editions, we plan to search and prove correct symbolic invariant conditions of the dynamics.

3.5 Lotka–Volterra with tangential crossings benchmark (LOVO21)

3.5.1 Model

The benchmark described below refers to the Lotka-Volterra equations, or predator-prey equations, which are well-known in the literature.

The system is defined as follows:

$$\begin{cases} \dot{x} = 3x - 3xy \\ \dot{y} = xy - y \end{cases} \tag{2}$$

which produces cyclic trajectories around the equilibrium point $(1, 1)$ dependent on the initial state.

We are interested to see how this nonlinear dynamics plays with a nonlinear guard, whose boundary is:

$$\sqrt{(x-1)^2 + (y-1)^2} = 0.161 \tag{3}$$

which is a circle of radius 0.161 around the equilibrium.

By choosing an initial state $I = (1.3, 1.0)$ the cycle has a period of approximately 3.64 time units. The trajectory of the Lotka–Volterra system trajectory is close to tangent to the guard circle in the top half, while it crosses the circle on the bottom half. Hence, enlarging the width of the initial set would put the trajectory partially within the guard in the top half.

The corresponding hybrid automaton is used to model the system:

- Continuous variables: x, y;
- Locations: *outside* and *inside*;

- Dynamics: those from Eq. 2 for x, y in both locations;
- Guards:
$$\begin{cases} (x-Q_x)^2 + (y-Q_y)^2 \leq R^2 \text{ from } \textit{outside} \text{ to } \textit{inside} \\ (x-Q_x)^2 + (y-Q_y)^2 \geq R^2 \text{ from } \textit{inside} \text{ to } \textit{outside} \end{cases} \tag{4}$$
- Invariants: the complement of the corresponding guards (i.e., transitions are urgent);
- Resets: none, i.e., the identity for both transitions.

3.5.2 Analysis

We want to start the system from $I = (1.3 \pm \epsilon, 1.0)$, with $\epsilon = 0.012$, and evolve it for $T = 3.64$ time units. Since the original system was close to tangency, by enlarging the initial set we expect to produce different sequences of discrete events due to the distinction between crossing and not crossing, and possibly by distinguishing the crossing sets based on the different crossing times. We must remark that, for reachability analysis purposes, it is important to carry the trace of discrete events along with the current evolution time.

The following three properties must be verified:

- At least one final set must have crossed two guards by entering and exiting the reference circle once;
- At least one final set must have crossed four guards by entering and exiting the reference circle twice;
- While a larger *even* number of crossings is allowed due to Zeno behavior during tangent crossing, no odd numbers are possible.

3.5.3 Evaluation

In terms of metrics, it is required to supply the following:

1. The execution time for computing the reachable set and checking the properties;
2. The area $x \times y$ of the box hull enclosing all the final sets.

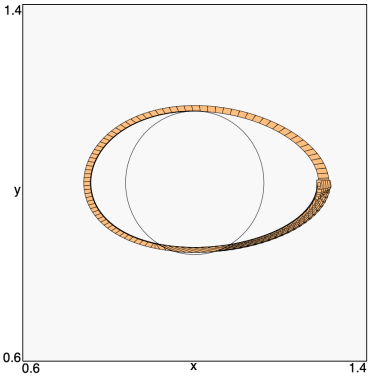
In addition, a figure showing the reachable set along with the circular guard shall be provided. The axes are $[0.6, 1.4] \times [0.6, 1.4]$.

3.5.4 Results

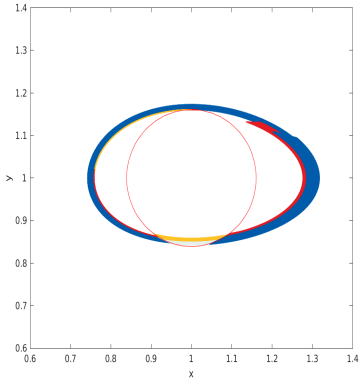
All tools were able to handle the benchmark with results equivalent to last year, except for a marked improvement in Kaa's set quality. Table 5 gives the timing/quality results, while Fig. 3 shows the graphical output.

Settings for Ariadne. A GradedTaylorSeriesIntegrator is used with a maximum spacial error of $1e-7$. The maximum step size is 0.07. The maximum number of parameters for a set is 5 times the number of variables, instead of the default of 3 times.

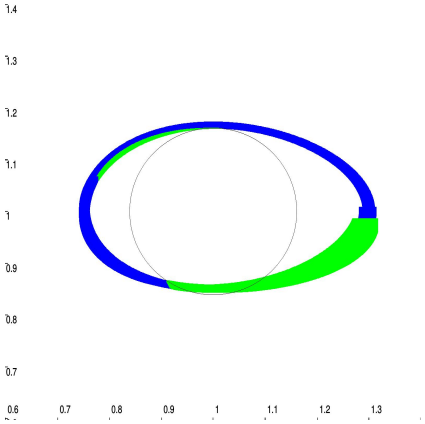
Settings for CORA. We use the approach in [33] to calculate the intersections with the nonlinear guard set. For continuous reachability we apply the conservative linearization approach [10] with time step size of 0.005 and a zonotope order of 20 for all modes.



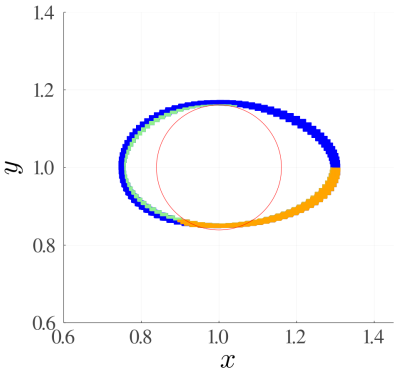
(a) Ariadne



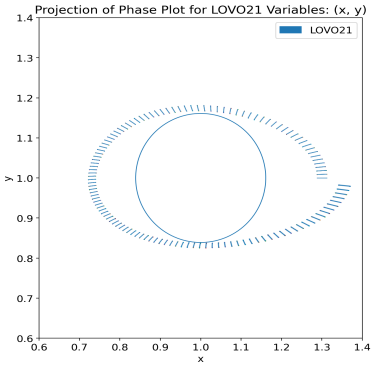
(b) CORA



(c) DynIbex



(d) JuliaReach



(e) Kaa

Figure 3: Reachable set overapproximation for LOVO21, with $x, y \in [0.6, 1.4]$, where the circular guard is shown.

Table 5: Results of LOVO21 in terms of computation time and area.

tool	**computation time in [s]**	**area**
Ariadne	11	1.1e-4
CORA	32	5.9e-3
DynIbex	75	5.9e-2
JuliaReach	3.8	1.4e-2
Kaa	568	4.2e-1
KeYmaera X	(0.4)[1]	0

[1] Duration of proving invariance (not checking crossing)

Settings for DynIbex. The library DynIbex does not support hybrid systems natively. However, based on constraint programming, event detection can be implemented and hybrid systems can be simulated. Reachability analysis is carried out with an error tolerance of 10^{-14} using an explicit Runge-Kutta method of order 4 (RK4 method). No splitting of the initial state has been performed.

Settings for Flow*. Since Flow* does not support urgent discrete transitions in hybrid systems, we skip the test on this benchmark.

Settings for Isabelle/HOL. Isabelle/HOL does not support hybrid systems automatically.

Settings for JuliaReach. We use $n_T = 7$, $n_Q = 1$, an adaptive absolute tolerance 10^{-10}, and split the initial set into 32 boxes. The crossings to the non-linear guard are handled by checking the reach sets that do not lie strictly outside the circle.

Settings for Kaa. Although Kaa does not support hybrid dynamics, we can plot the reachable set utilizing a time step of $\Delta = 0.03$ for the dynamics given above. This year we changed the dynamic template parameters to have PCA templates with a lifespan of 10 steps as well as linear approximation templates with a lifespan of 10 steps. Although the volume of the reachable set has diminished from the previous year, the running time has increased in exchange.

Settings for KeYmaera X. The KeYmaera X proof focuses on infinite-horizon population stability for any positive starting choice of populations $x > 0$ and $y > 0$, which includes the specific starting populations $x = 1.3 \pm \epsilon$ and $y = 1$. The population orbit is stable around $(\frac{\alpha}{\beta}, \frac{\gamma}{\delta})$ at population $e^{-\delta x - \beta y} x^\gamma y^\alpha$ for $\alpha = \beta = 3$ and $\delta = \gamma = 1$.

```
1  Definitions Real K(Real x, Real y) = exp(−d∗x−b∗y) ∗ x^g ∗ y^a; End.
2  Problem
3    a=3 & b=3 & d=1 & g=1 & x>0 & y>0 & K_0 = K(x,y)
4    −>
5    [{ x' = a∗x − b∗x∗y,
6       y' = d∗x∗y − g∗y
7     }
8    ]K(x,y) = K_0
9  End.
10
11 Tactic "Scripted proof"
12 unfold;
13 dIRule(1); <(
```

```
14    "dI Init": equalCommute(1); id,
15    "dI Step":
16      chaseAt(1);
17      QE using "(exp(−1∗x−3∗y)∗(−1∗(3∗x−3∗x∗y)−3∗(1∗x∗y−1∗y))∗x^1+exp(−1∗x−3∗y)∗(1∗x^(1−1)∗(3∗x
            ↪ −3∗x∗y)))∗y^3+exp(−1∗x−3∗y)∗x^1∗(3∗y^(3−1)∗(1∗x∗y−1∗y))=0"
18  )
19  End.
20  Tactic "Automated proof" autoClose End.
```

The formalization in the repeatability package also includes a symbolic characterization of the existence of crossing in and out of the nonlinear guard: this purely real arithmetic proof obligation is not yet tractable by the arithmetic backend verification procedures used in KeYmaera X. In future editions, we plan to additionally characterize the number of transitions symbolically.

3.6 Space rendezvous benchmark (SPRE22)

3.6.1 Model

Spacecraft rendezvous is a perfect use case for formal verification of hybrid systems with nonlinear dynamics since mission failure can cost lives and is extremely expensive. This benchmark is taken from [22]. A version of this benchmark with linearized dynamics is verified in the ARCH-COMP category *Continuous and Hybrid Systems with Linear Continuous Dynamics*. The nonlinear dynamic equations describe the two-dimensional, planar motion of the spacecraft on an orbital plane towards a space station:

$$\begin{cases} \dot{x} &= v_x \\ \dot{y} &= v_y \\ \dot{v_x} &= n^2 x + 2nv_y + \frac{\mu}{r^2} - \frac{\mu}{r_c^3}(r+x) + \frac{u_x}{m_c} \\ \dot{v_y} &= n^2 y - 2nv_x - \frac{\mu}{r_c^3} y + \frac{u_y}{m_c} \end{cases}$$

The model consists of position (relative to the target) x, y [m], time t [min], as well as horizontal and vertical velocity v_x, v_y [m / min]. The parameters are $\mu = 3.986 \times 10^{14} \times 60^2$ [m^3 / min^2], $r = 42164 \times 10^3$ [m], $m_c = 500$ [kg], $n = \sqrt{\frac{\mu}{r^3}}$ and $r_c = \sqrt{(r+x)^2 + y^2}$.

The hybrid nature of this benchmark originates from a switched controller. In particular, the modes are *approaching* ($x \in [-1000, -100]$ [m]), *rendezvous attempt* ($x \geq -100$ [m]), and *aborting*. A transition to mode *aborting* occurs nondeterministically at $t \in [120, 150]$ [*min*]. The linear feedback controllers for the different modes are defined as $\binom{u_x}{u_y} = K_1 \underline{x}$ for mode *approaching*, and $\binom{u_x}{u_y} = K_2 \underline{x}$ for mode *rendezvous attempt*, where $\underline{x} = \begin{pmatrix} x & y & v_x & v_y \end{pmatrix}^T$ is the vector of system states. The feedback matrices K_i were determined with an LQR-approach applied to the linearized system dynamics, which resulted in the following numerical values:

$$K_1 = \begin{pmatrix} -28.8287 & 0.1005 & -1449.9754 & 0.0046 \\ -0.087 & -33.2562 & 0.00462 & -1451.5013 \end{pmatrix}$$

$$K_2 = \begin{pmatrix} -288.0288 & 0.1312 & -9614.9898 & 0 \\ -0.1312 & -288 & 0 & -9614.9883 \end{pmatrix}$$

In the mode *aborting*, the system is uncontrolled $\binom{u_x}{u_y} = \binom{0}{0}$.

Table 6: Results of SPRE21 in terms of computation time.

tool	**computation time in [s]**
Ariadne	–
CORA	42
DynIbex	61
JuliaReach	21
Kaa	N/A
KeYmaera X	N/A

3.6.2 Analysis

The spacecraft starts from the initial set $x \in [-925, -875]$ [m], $y \in [-425, -375]$ [m], $v_x \in [0, 5]$ [m/min] and $v_y \in [0, 5]$ [m/min]. For the considered time horizon of $t \in [0, 200]$ [min], the following specifications have to be satisfied:

- **Line-of-sight:** In mode *rendezvous attempt*, the spacecraft has to stay inside line-of-sight cone $\mathcal{L} = \{\binom{x}{y} \mid (x \geq -100) \wedge (y \geq x \tan(20°)) \wedge (-y \geq x \tan(20°))\}$.
- **Collision avoidance:** In mode *aborting*, the spacecraft has to avoid a collision with the target, which is modeled as a box $\mathcal{B}$ with 2m edge length and the center placed at the origin.
- **Velocity constraint:** In mode *rendezvous attempt*, the absolute velocity has to stay below 3.3 [m/min]: $\sqrt{v_x^2 + v_y^2} \leq 3.3$ [m/min].

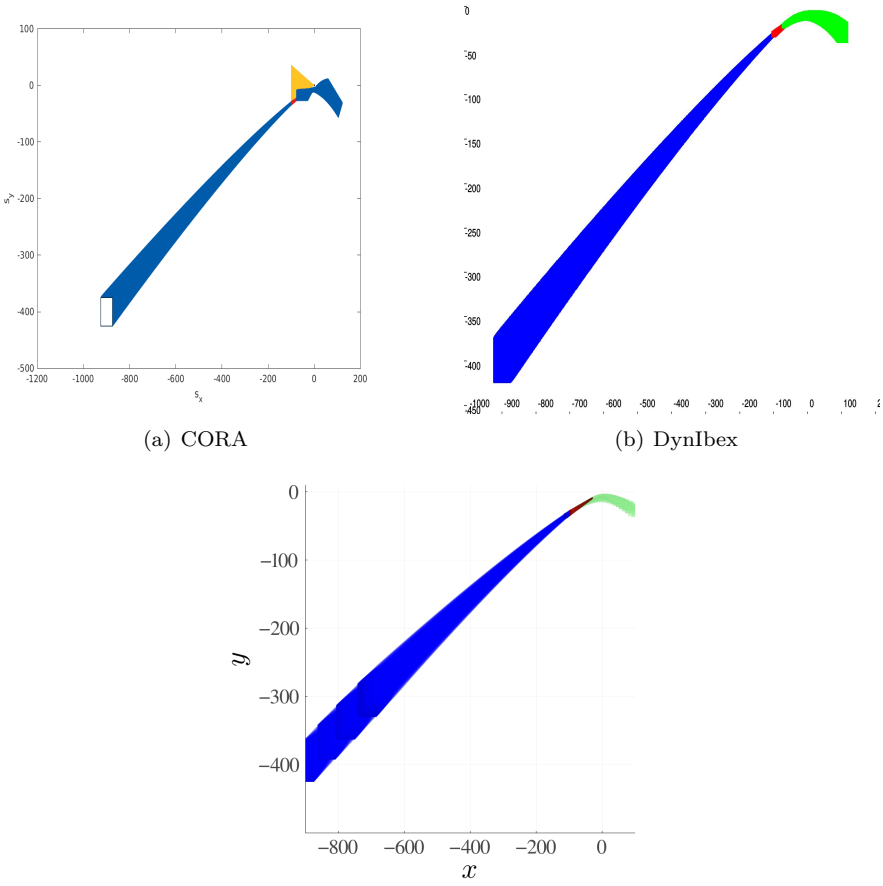
Remark on velocity constraint In the original benchmark [22], the constraint on the velocity was set to 0.05 m/s, but it can be shown (by a counterexample) that this constraint cannot be satisfied. We therefore use the relaxed constraint 0.055 [m/s] = 3.3 [m/min].

3.6.3 Evaluation

The computation time for evolution and verification is provided. A figure is shown in the (x, y) axes, with $x \in [-1000, 200]$ and $y \in [-450, 0]$.

3.6.4 Results

The results of the reachability computation for the spacecraft rendezvous model are given in Figure 4 and Table 6, with the tool settings below. The introduction of a permissive guard prevented completion for Ariadne: too many trajectories were generated and the absence of a recombination strategy proved an issue. Therefore this benchmark requires proper support of crossings in the presence of large sets, even if the crossing region is very simple from a geometrical viewpoint. The hybrid nature of the problem was an obstacle for Kaa. KeYmaera X formalized but not proved the problem yet.



(a) CORA

(b) DynIbex

(c) JuliaReach

Figure 4: Reachable set of the spacecraft position in the x-y-plane for SPRE21.

Settings for Ariadne. Ariadne was not able to complete evolution, due to the extremely large number of trajectories produced from the nondeterministic guard: this is caused by the lack of a recombination strategy. The maximum step size used was 1.0, essentially meaning that we allowed the step size to vary widely along evolution: this choice turned out to be preferable in terms of execution time. The maximum temporal order was 4 and the maximum spacial error enforced for each step equal is 10^{-3}. A splitting strategy for the initial set was used; the strategy compare the radius of the set with a reference value of 12.0, in order to split the first two dimensions once and yield a total of 4 initial subsets.

Settings for CORA. CORA was run with a time step size of 0.2 [min] for the modes *approaching* and *aborting*, and with a time step size of 0.05 [min] for mode *rendezvous attempt*. The intersections with the guard sets are calculated with constrained zonotopes [44], and the intersection is then enclosed with a zonotope bundle [9]. In order to find suitable orthogonal directions for the enclosure *principal component analysis* is applied.

Settings for DynIbex. The library DynIbex does not support hybrid systems natively. However, based on constraint programming, event detection can be implemented and hybrid systems can be simulated. Maximum zonotope order is set to 10, reachability analysis is carried out with an error tolerance of 10^{-6} using an explicit Runge-Kutta method of order 3 (Kutta's method). No splitting of the initial state has been performed.

Settings for JuliaReach. The transition to the aborting mode is handled by clustering and Cartesian decomposition [19] with zonotope enclosures in low dimensions, (x, y) and (v_x, v_y). The continuous-time algorithms used in the modes approaching, rendez-vous attempt, and aborting are `TMJets20` (first two modes) and `TMJets21b` (third mode) with $n_T = 5, 4, 7$, $n_Q = 1, 1, 1$ and adaptive absolute tolerance $10^{-5}, 10^{-7}, 10^{-10}$, respectively.

Settings for Kaa. Kaa does not support hybrid dynamics in its current state.

Settings for KeYmaera X. The example was formalized for KeYmaera X but not yet proved. The full model is included in the repeatability evaluation package.

4 Conclusion and Outlook

This year, the competition confirmed the five participants from 2021 and added KeYmaera X as a new entry. Given the theorem proving nature of this last tool, it was interesting to see how it would behave as opposed to the remaining numerical competitors. KeYmaera X showed remarkable performances, although it was not possible to tackle all the benchmarks: hence, we will have to wait next year for a full evaluation.

Speaking about benchmark evaluation, this year we chose not to drop any benchmark and rather accept that some tools may not be able to address the whole suite. As such, we modified two benchmarks, namely van der Pol (CVDP22) and Space Rendezvous (SPRE22) and added a new one (Traffic scenario, TRAF22).

The CVDP22 benchmark proved a bit too difficult, since only the proponent (JuliaReach's team) was able to solve it. For the next year, we will definitely have to simplify the problem to accommodate more tools.

The SPRE22 benchmark instead was solved by those tools that could solve the previous year's instance. Other tools either have numerical issues (Ariadne), can't express hybrid dynamics (Kaa) or were not able to prove the problem for this year yet (KeYmaera X).

The benchmarks that remained the same as 2021 were ROBE21, LALO20 and LOVO21, for which we did not notice particular improvements except for Kaa on LOVO21.

We care to mention that, triggered by the participation in this competition, individual tools made progress:

- The algorithms for nonlinear continuous-time systems used in CORA have been extended by adaptive parameter tuning. However, the benchmarks in this competition are too close to the computational limits of the reachability algorithms so that manual tuning is still required. We expect to make more use of adaptive parameter tuning in future editions.

- JuliaReach used the same algorithms as last year and produced the same qualitative results. Still, there were several improvements in the following core packages: `ReachabilityAnalysis.jl`, `LazySets.jl` and `TaylorModels.jl`, which sum up to minor quantitative improvements.

Summarizing, the new benchmark and their variations were interesting by way of continuing to explore critical aspects of continuous/hybrid evolution, with the objective of pushing all tools forward. We believe that a benchmark suite with representative problems is of the utmost importance, in order to stimulate meaningful progress of all the participating tools. At the same time, we care about allowing all tools to solve all benchmarks and we will try to modify the most critical ones in order to achieve that. Consequently, for the next year we aim at refining the existing suite to advance in these directions, also possibly increasing the number of benchmarks.

5 Acknowledgments

Luca Geretti acknowledges the support from MIUR, Project "Dipartimenti d'Eccellenza, 2018-2022" and by Fondazione Cariverona with the grant "Ricerca&Sviluppo".

Luis Benet acknowledges support from PAPIIT grant IG-101122.

Matthias Althoff acknowledges support by the European Commission project justITSELF under grant number 817629.

Mark Wetzlinger acknowledges support from the research training group CONVEY funded by the German Research Foundation under grant GRK 2428.

Julien Alexandre dit Sandretto: this work was supported by the "Chair Complex Systems Engineering - Ecole polytechnique, THALES, DGA, FX, Dassault Aviation, DCNS Research, ENSTA Paris, Télécom Paris, and Fondation ParisTech" and partially supported by DGA AID.

Parasara Sridhar Duggirala and Edward Kim acknowledge the support of the Air Force Office of Scientific Research under award number FA9550-19-1-0288 and FA9550-21-1-0121 and National Science Foundation (NSF) under grant numbers CNS 1935724 and CNS 2038960. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the United States Air Force or National Science Foundation.

Stefan Mitsch was sponsored by the AFOSR under grant number FA9550-16-1-0288. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of any sponsoring institution, the U.S. government or any other entity.

Christian Schilling acknowledges support from DIREC - Digital Research Centre Denmark and the Villum Investigator Grant S4OS.

References

[1] Kodiak, a C++ library for rigorous branch and bound computation. https://github.com/nasa/Kodiak, Accessed: July 2020.

[2] Julien Alexandre dit Sandretto and Alexandre Chapoutot. Validated Explicit and Implicit Runge-Kutta Methods. *Reliable Computing electronic edition*, 22, 2016.

[3] Julien Alexandre dit Sandretto and Alexandre Chapoutot. Validated Simulation of Differential Algebraic Equations with Runge-Kutta Methods. *Reliable Computing electronic edition*, 22, 2016.

[4] Julien Alexandre Dit Sandretto, Alexandre Chapoutot, and Olivier Mullier. Constraint-Based Framework for Reasoning with Differential Equations. In Çetin Kaya Koç, editor, *Cyber-Physical Systems Security*, pages 23–41. Springer International Publishing, December 2018.

[5] M. Althoff. Reachability analysis of nonlinear systems using conservative polynomialization and non-convex sets. In *Hybrid Systems: Computation and Control*, pages 173–182, 2013.

[6] M. Althoff. An introduction to CORA 2015. In *Proc. of the Workshop on Applied Verification for Continuous and Hybrid Systems*, pages 120–151, 2015.

[7] M. Althoff and D. Grebenyuk. Implementation of interval arithmetic in CORA 2016. In *Proc. of the 3rd International Workshop on Applied Verification for Continuous and Hybrid Systems*, pages 91–105, 2016.

[8] M. Althoff, M. Koschi, and S. Manzinger. Commonroad: Composable benchmarks for motion planning on roads. In *Proc. of the IEEE Intelligent Vehicles Symposium*, 2017.

[9] M. Althoff and B. H. Krogh. Zonotope bundles for the efficient computation of reachable sets. In *Proc. of the 50th IEEE Conference on Decision and Control*, pages 6814–6821, 2011.

[10] M. Althoff, O. Stursberg, and M. Buss. Reachability analysis of nonlinear systems with uncertain parameters using conservative linearization. In *Proc. of the 47th IEEE Conference on Decision and Control*, pages 4042–4048, 2008.

[11] Miguel Angel Barron. Stability of a ring of coupled van der Pol oscillators with non-uniform distribution of the coupling parameter. In *Journal of applied research and technology 14.1*, pages 62–66, 2016.

[12] Luis Benet and David P. Sanders. TaylorSeries.jl: Taylor expansions in one and several variables in Julia. *Journal of Open Source Software*, 4(36):1043, April 2019. doi:10.21105/joss.01043.

[13] Luis Benet and David P. Sanders. JuliaDiff/TaylorSeries.jl. https://github.com/JuliaDiff/TaylorSeries.jl, April 2021. doi:10.5281/zenodo.2601941.

[14] Luis Benet and David P. Sanders. JuliaIntervals/IntervalArithmetic.jl. https://github.com/JuliaIntervals/IntervalArithmetic.jl, May 2021. doi:10.5281/zenodo.3336308.

[15] Luis Benet and David P. Sanders. JuliaIntervals/TaylorModels.jl. https://github.com/JuliaIntervals/TaylorModels.jl, June 2021. doi:10.5281/zenodo.2613102.

[16] L. Benvenuti, D. Bresolin, P. Collins, A. Ferrari, L. Geretti, and T. Villa. Assume-guarantee verification of nonlinear hybrid systems with Ariadne. *Int. J. Robust. Nonlinear Control*, 24(4):699–724, 2014.

[17] Carl Boettiger. An introduction to docker for reproducible research. *ACM SIGOPS Operating Systems Review*, 49(1):71–79, 2015.

[18] S. Bogomolov, M. Forets, G. Frehse, K. Potomkin, and C. Schilling. JuliaReach: a toolbox for set-based reachability. In *HSCC*, 2019. doi:10.1145/3302504.3311804.

[19] Sergiy Bogomolov, Marcelo Forets, Goran Frehse, Andreas Podelski, and Christian Schilling. Decomposing reach set computations with low-dimensional sets and high-dimensional matrices. *Inf. Comput.*, 2022. doi:10.1016/j.ic.2022.104937.

[20] Davide Bresolin, Pieter Collins, Luca Geretti, Roberto Segala, Tiziano Villa, and Sanja Živanović Gonzalez. A Computable and Compositional Semantics for Hybrid Automata. In *Proceedings of the 23rd International Conference on Hybrid Systems: Computation and Control*, HSCC '20, New York, NY, USA, 2020. Association for Computing Machinery. doi:10.1145/3365365.3382202.

[21] Davide Bresolin, Luca Geretti, Riccardo Muradore, Paolo Fiorini, and Tiziano Villa. Formal verification applied to robotic surgery. *Lecture Notes in Control and Information Sciences*, 456:347 – 355, 2015. doi:10.1007/978-3-319-10407-2_40.

[22] N. Chan and S. Mitra. Verifying safety of an autonomous spacecraft rendezvous mission. In *ARCH17. 4th International Workshop on Applied Verification of Continuous and Hybrid Systems, collocated with Cyber-Physical Systems Week (CPSWeek) on April 17, 2017 in Pittsburgh, PA, USA*, pages 20–32, 2017. URL: http://www.easychair.org/publications/paper/342723.

[23] P. Collins, D. Bresolin, L. Geretti, and T. Villa. Computing the evolution of hybrid systems using rigorous function calculus. In *Proc. of the 4th IFAC Conference on Analysis and Design of Hybrid Systems (ADHS12)*, pages 284–290, Eindhoven, The Netherlands, June 2012.

[24] Tommaso Dreossi. Sapo: Reachability computation and parameter synthesis of polynomial dynamical systems. In *Proceedings of the 20th International Conference on Hybrid Systems: Computation and Control*, pages 29–34, 2017.

[25] Tommaso Dreossi, Thao Dang, and Carla Piazza. Parallelotope bundles for polynomial reachability. In *Proceedings of the 19th International Conference on Hybrid Systems: Computation and Control*, pages 297–306, 2016.

[26] Vincent Drevelle and Jeremy Nicola. Vibes: A visualizer for intervals and boxes. *Mathematics in Computer Science*, 8(3):563–572, Sep 2014.

[27] Marcelo Forets and Christian Schilling. LazySets.jl: Scalable symbolic-numeric set computations. *Proceedings of the JuliaCon Conferences*, 1(1):11, 2021. doi:10.21105/jcon.00097.

[28] Nathan Fulton, Stefan Mitsch, Rose Bohrer, and André Platzer. Bellerophon: Tactical theorem proving for hybrid systems. In *Interactive Theorem Proving - 8th International Conference, ITP 2017, Brasília, Brazil, September 26-29, 2017, Proceedings*, pages 207–224, 2017. doi:10.1007/978-3-319-66107-0_14.

[29] Nathan Fulton, Stefan Mitsch, Jan-David Quesel, Marcus Völp, and André Platzer. Keymaera X: an axiomatic tactical theorem prover for hybrid systems. In *Automated Deduction - CADE-25 - 25th International Conference on Automated Deduction, Berlin, Germany, August 1-7, 2015, Proceedings*, pages 527–538, 2015. doi:10.1007/978-3-319-21401-6_36.

[30] James Gallicchio, Yong Kiam Tan, Stefan Mitsch, and André Platzer. Implicit definitions with differential equations for keymaera x (system description). In Jasmin Blanchette, Laura Kovacs, and Dirk Pattinson, editors, *IJCAR*, volume 13385 of *LNCS*. Springer, 2022. doi:10.1007/978-3-031-10769-6_42.

[31] Edward Kim, Stanley Bak, and Parasara Sridhar Duggirala. Automatic dynamic parallelotope bundles for reachability analysis of nonlinear systems, 2021. arXiv:2105.11796.

[32] Edward Kim and Parasara Sridhar Duggirala. Kaa: A python implementation of reachable set computation using bernstein polynomials. *EPiC Series in Computing*, 74:184–196, 2020.

[33] N. Kochdumper and M. Althoff. Reachability analysis for hybrid systems with nonlinear guard sets. In *Proceedings of the 23rd International Conference on Hybrid Systems: Computation and Control*, 2020.

[34] N. Kochdumper, P. Gassert, and M. Althoff. Verification of collision avoidance for CommonRoad traffic scenarios. In *Proc. of the 8th International Workshop on Applied Verification of Continuous and Hybrid Systems*, pages 184–194, 2021. doi:10.29007/1973.

[35] M. T. Laub and W. F. Loomis. A molecular network that produces spontaneous oscillations in excitable cells of dictyostelium. *Molecular Biology of the Cell*, 9:3521–3532, 1998.

[36] Stefan Mitsch. Implicit and explicit proof management in keymaera X. In *Proceedings of the 6th Workshop on Formal Integrated Development Environment, F-IDE@NFM 2021, Held online, 24-25th May 2021*, pages 53–67, 2021. doi:10.4204/EPTCS.338.8.

[37] Stefan Mitsch and André Platzer. The keymaera X proof IDE - concepts on usability in hybrid systems theorem proving. In *Proceedings of the Third Workshop on Formal Integrated Development Environment, F-IDE@FM 2016, Limassol, Cyprus, November 8, 2016*, pages 67–81, 2016. doi:10.4204/EPTCS.240.5.

[38] Stefan Mitsch and André Platzer. A retrospective on developing hybrid system provers in the keymaera family - A tale of three provers. In *Deductive Software Verification: Future Perspectives - Reflections on the Occasion of 20 Years of KeY*, pages 21–64. 2020. doi:10.1007/978-3-030-64354-6_2.

[39] Olivier Mullier, Alexandre Chapoutot, and Julien Alexandre dit Sandretto. Validated computation of the local truncation error of runge–kutta methods with automatic differentiation. *Optimization Methods and Software*, 33(4-6):718–728, 2018.

[40] Jorge A. Pérez-Hernández and Luis Benet. PerezHz/TaylorIntegration.jl. https://github.com/PerezHz/TaylorIntegration.jl, May 2021. doi:10.5281/zenodo.2562352.

[41] André Platzer. A complete uniform substitution calculus for differential dynamic logic. *J. Autom. Reason.*, 59(2):219–265, 2017. doi:10.1007/s10817-016-9385-1.

[42] André Platzer and Yong Kiam Tan. Differential equation invariance axiomatization. *J. ACM*, 67(1):6:1–6:66, 2020. doi:10.1145/3380825.

[43] H. H. Robertson. The solution of a set of reaction rate equations. In *"Numerical analysis: an introduction"*, page 178–182. Academic Press, 1966.

[44] J. K. Scott, D. M. Raimondo, G. R. Marseglia, and R. D. Braatz. Constrained zonotopes: A new tool for set-based estimation and fault detection. *Automatica*, 69:126–136, 2016.

[45] Andrew Sogokon, Stefan Mitsch, Yong Kiam Tan, Katherine Cordwell, and André Platzer. Pegasus: sound continuous invariant generation. *Formal Methods Syst. Des.*, 58(1-2):5–41, 2021. doi:10.1007/s10703-020-00355-z.

[46] Yong Kiam Tan, Stefan Mitsch, and André Platzer. Verifying switched system stability with logic. In *HSCC '22: 25th ACM International Conference on Hybrid Systems: Computation and Control, Milan, Italy, May 4 - 6, 2022*, pages 2:1–2:11, 2022. doi:10.1145/3501710.3519541.

[47] R. Testylier and T. Dang. Nltoolbox: A library for reachability computation of nonlinear dynamical systems. In *Proc. of ATVA'13*, volume 8172 of *LNCS*, pages 469–473. Springer, 2013.

[48] K. Weihrauch. *Computable analysis*. Texts in Theoretical Computer Science. An EATCS Series. Springer-Verlag, Berlin, 2000.