

The 3rd Reactive Synthesis Competition (SYNTCOMP 2016): Benchmarks, Participants & Results

Swen Jacobs Saarland University Saarbrücken, Germany	Roderick Bloem Graz University of Technology Graz, Austria	Romain Brenguier University of Oxford Oxford, UK	
Ayrat Khalimov Graz University of Technology Graz, Austria	Felix Klein Saarland University Saarbrücken, Germany	Robert Könighofer Graz University of Technology Graz, Austria	
Jens Kreber Saarland University Saarbrücken, Germany	Alexander Legg Data 61, CSIRO (formerly NICTA) and UNSW Sydney, Australia	Nina Narodytska Samsung Research America Mountain View, USA	
Guillermo A. Pérez Université Libre de Bruxelles Brussels, Belgium	Jean-François Raskin Université Libre de Bruxelles Brussels, Belgium	Leonid Ryzhyk Samsung Research America Mountain View, USA	
Ocan Sankur CNRS, Irisa Rennes, France	Martina Seidl Johannes-Kepler-University Linz, Austria	Leander Tentrup Saarland University Saarbrücken, Germany	Adam Walker Independent Researcher

We report on the benchmarks, participants and results of the third reactive synthesis competition (SYNTCOMP 2016). The benchmark library of SYNTCOMP 2016 has been extended to benchmarks in the new LTL-based *temporal logic synthesis format* (TLSF), and 2 new sets of benchmarks for the existing AIGER-based format for safety specifications. The participants of SYNTCOMP 2016 can be separated according to these two classes of specifications, and we give an overview of the 6 tools that entered the competition in the AIGER-based track, and the 3 participants that entered the TLSF-based track. We briefly describe the benchmark selection, evaluation scheme and the experimental setup of SYNTCOMP 2016. Finally, we present and analyze the results of our experimental evaluation, including a comparison to participants of previous competitions and a legacy tool.

1 Introduction

Since the definition of the problem more than 50 years ago [16], the automatic synthesis of reactive systems from formal specifications is one of the major challenges of computer science. Research into the basic questions related to the problem has led to a large body of theoretical results, but their impact on the practice of system design has been rather limited. To increase the impact of theoretical advancements in synthesis, the reactive synthesis competition (SYNTCOMP) has been founded in 2014 [27]. The competition is designed to foster research in scalable and user-friendly implementations of synthesis techniques by establishing a standard benchmark format, maintaining a challenging public benchmark library, and providing a *dedicated and independent* platform for the comparison of tools under consistent experimental conditions.

The first SYNTCOMP was held in 2014 [27], and the second in 2015 [28]. The competition is regularly associated with the International Conference on Computer Aided Verification (CAV) and the Workshop on Synthesis (SYNT), and the results of the annual competitions are first presented at these

venues. A design choice for the first two competitions was to focus on safety properties specified as monitor circuits in an extension of the AIGER format known from the hardware model checking competition [15, 25]. SYNTCOMP 2016 introduces the first major extension of the competition: in addition to the existing competition track, we introduce a new track that is based on properties in full linear temporal logic (LTL), given in the *temporal logic synthesis format* (TLSF) recently introduced by Jacobs, Klein and Schirmer [29].

The organization team of SYNTCOMP 2016 consisted of R. Bloem and S. Jacobs, with technical assistance from J. Kreber for the setup and execution, and from F. Klein for the integration of TLSF.

The rest of this paper describes the design, benchmarks, participants, and results of SYNTCOMP 2016. We present the benchmark set for SYNTCOMP 2016 in Section 2, followed by a description of the setup, rules and execution of the competition in Section 3. In Section 4 we give an overview of the participants of SYNTCOMP 2016, focusing on changes compared to last year’s participants. Finally, the experimental results are presented and analyzed in Section 5.

Note that more details on the goals and design of the competition can be found in the sister paper that discusses the design of SYNTCOMP in 2016 and the future [26].

2 Benchmarks

In this section, we describe the benchmark library for SYNTCOMP 2016. We start by describing benchmarks in the new TLSF format, for the LTL synthesis track. Then, we describe new benchmarks in the AIGER format, followed by a summary of the classes of benchmarks in AIGER format that have already been used in SYNTCOMP 2014 and 2015.

2.1 TLSF Benchmarks

All of the following benchmarks have been translated into TLSF or directly encoded in TLSF by F. Klein, except for the last set of benchmarks, that has been encoded by S. Jacobs.

Lily benchmark set. This set of benchmarks was originally included with the LTL synthesis tool LILY¹ [32]. It includes 24 benchmarks.

Acacia benchmark set. This set of benchmarks was originally included with the LTL synthesis tool Acacia+² [9]. It includes 65 benchmarks.

Parameterized detector. This benchmark specifies a simple component that gets a number of inputs and raises its single output infinitely often if and only if all of the inputs are raised infinitely often. Parameterized in the number of inputs.

Parameterized arbiters. This set of benchmarks includes 4 different arbiter specifications, where a central controller receives requests and gives (mutually exclusive) grants to a number of communication

¹LILY is available at http://www.iaik.tugraz.at/content/research/design_verification/lily/. Accessed August 2016.

²Acacia+ is available at <http://lit2.ulb.ac.be/acaciaplus/>. Accessed August 2016.

partners, called *masters*. The basic specification is the *simple arbiter*, which only requires mutual exclusion and request-response. Additionally, there are 3 advanced specifications that require additional properties: *prioritized arbiter*, *round-robin arbiter*, and *full arbiter* (including absence of spurious grants). All benchmarks are parameterized in the number of masters n that the arbiter needs to serve.

Parameterized AMBA bus controller. The AMBA bus controller is essentially a very advanced arbiter that includes additional features, like locking the bus for a fixed or arbitrary number of steps. It has been translated from an industrial specification in natural language³ to LTL by Jobstmann [31], and is one of the most challenging and widely used benchmark for synthesis tools [5, 8, 20, 24]. Like the arbiters, the benchmark is parameterized in the number of masters that the controller has to serve.

Parameterized load balancer. This benchmark considers a load balancer that receives jobs and distributes them to a fixed number of servers, with a number of additional properties like prioritization of the first server, and that the load is balanced between servers. The original specification has been designed by Rüdiger Ehlers [18] for the evaluation of synthesis tool UNBEAST [17]. The benchmark is parameterized in the number of servers that can handle the jobs.

Parameterized generalized buffer. This benchmark specifies a family of buffers that transmit data from a number of senders to two receivers, based on a handshake protocol between the buffer and the other components, and an interface to a FIFO queue that is used to store data. The benchmark is parameterized in the number of senders. Like the AMBA case study, it has first been translated to LTL by Jobstmann [31].

Parameterized LTL to Büchi translation (LTL2DBA). This set of benchmarks uses the synthesis procedure to generate deterministic Büchi automata that correspond to an LTL specification. It is based on a number of parameterized LTL formulas, as described by Tian et al. [47].

2.2 AIGER Benchmarks

HWMCC benchmarks. This class of benchmarks is based on a subset of the benchmarks from HWMCC 2012 [15]. The idea is to consider verification benchmarks consisting of a system and a safety specification that is not satisfied by the system, and ask the question whether we can synthesize a controller for a given subset of the inputs such that the specification is satisfied.

Benchmarks in this set are based on the unsafe instances from the *single safety property* track of HWMCC 2012, modified such that between 1 and 8 inputs of the circuit are defined to be controllable. The files retain their original filenames, with `_c0toCx` appended to the filename if inputs 0 to x are declared as controllable. E.g., the benchmark based on `6s5.aag`, where 8 inputs have been declared as controllable, is named `6s5_c0to7.aag`.

This benchmark set contains 280 benchmarks. The original verification benchmarks have been taken from the HWMCC website⁴ and modified by S. Jacobs.

³AMBA Specification Rev 2.0, available at <http://www.arm.com/>. Accessed August 2016.

⁴HWMCC website: <http://fmv.jku.at/hwmcc/>. Accessed August 2016.

(Bounded) LTL to Büchi and LTL to parity translation (LTL2DBA/LTL2DPA). This class of benchmarks is based on the benchmarks for LTL to Büchi translation mentioned in Section 2.1, but for some instances also considers the synthesis of parity automata. In addition to the original parameterization, they are parameterized in the liveness-to-safety approximation. This set contains 62 instances and have been translated from TLSF to AIGER by G. A. Pérez.

Existing benchmark library. The existing library that was the basis for SYNTCOMP 2014 and 2015 contains 3105 benchmark instances. The following list of benchmarks was collected for the first competition, and more details can be found in the SYNTCOMP 2014 report [27]:

- Toy Examples: this benchmark set contains specifications for a number of basic building blocks of circuits, such as an adder (`add`), a bitshifter (`bs`), a counter (`count`), or a multiplier (`mult`). All of the benchmarks are parameterized in the bitwidth of the input, resulting in 176 problem instances.
- AMBA: a version of the bus controller specification for AMBA (as described by Bloem et al. [5]), parameterized in three dimensions (number of masters, and the type and the precision of the liveness-to-safety approximation). The benchmark set contains 952 instances.
- Genbuf: a version of the generalized buffer specification (as described by Bloem et al. [5]), parameterized in the same way as the AMBA benchmarks above. The set contains 866 instances.
- LTL2AIG: these are several sets of benchmarks that are based on the benchmark set of synthesis tool Acacia+ [9], translated using the LTL2AIG tool [27]:
 - 50 demo problems originally supplied with LTL synthesis tool LILY,
 - 41 `ltl2dba` and `ltl2dpa` problems that use the synthesis tool to obtain a deterministic Büchi automaton (`dba`) or a deterministic parity automaton (`dpa`) that corresponds to a given LTL formula,
 - 42 `gb` benchmarks, based on the generalized buffer specification, but with a notably higher difficulty than the version mentioned above,
 - 64 load balancer benchmarks, a modification of those originally presented with LTL synthesis tool UNBEAST⁵ [17].
- Factory Assembly Line: this benchmark set describes a controller for two robot arms on an assembly line. It consists of 15 problem instances.
- Moving Obstacle: this benchmark set describes a moving robot that should avoid a moving obstacle in two-dimensional space. It consists of 16 problem instances.

While all classes above have been introduced for SYNTCOMP 2014, some of the instances have only been generated for SYNTCOMP 2015. Additionally, the following classes have been added for the second competition, and have been described in the SYNTCOMP 2015 report [28]:

- Washing Cycle Scheduler: specifications for a controller of a washing system, with water tanks that share pipes and cycles that can be launched in parallel. Parameterized in the number of tanks, the maximum reaction delay, and the shared water pipes. This benchmark set contains 321 instances.
- Driver Synthesis: this class of benchmarks specifies a driver for a hard disk controller with respect to a given operating system model (as described by Ryzhyk et al. [43]). It is parameterized in the level of data abstraction, the precision of the liveness-to-safety approximation, and the simplification of the specification circuit by ABC [11]. This benchmark set contains 72 instances.

⁵UNBEAST is available at <http://www.react.uni-saarland.de/tools/unbeast/>. Accessed August 2016.

- **Huffman Encoder:** this class of benchmarks specifies a given Huffman decoder, for which a suitable encoder should be synthesized (as described by Khalimov [33]). The benchmark is parameterized in the liveness-to-safety approximation, resulting in 5 problem instances.
- **HWMCC:** this class of benchmarks is based on a subset of the benchmarks from HWMCC 2014 [15], modified in the same way as the new HWMCC benchmarks mentioned above. This set contains 110 instances.
- **HyperLTL:** this class of benchmarks is based on benchmark problems from HyperLTL model checking, as introduced in recent work by Finkbeiner et al. [21]. The goal is to synthesize a witness formula (represented as a controller circuit) for a given HyperLTL property. This benchmark set contains 21 instances.
- **Matrix Multiplication:** these benchmarks specify matrix multiplication circuits, parameterized in the size of the input matrices. The basic benchmarks consider a single multiplication, and a variant models repeated multiplication with a subset of controllable inputs and an additional safety goal. This benchmark set contains 273 instances for basic case, and 81 for the repeated case.

3 Setup, Rules and Execution

We give an overview of the setup, rules and execution of SYNTCOMP 2016. More details, as well as information about the reasoning behind different design choices, can be found in the sister paper by Jacobs and Bloem [26].

3.1 General Rules

Like in the previous year, there is a track that is based on safety specifications in AIGER format (in the following: AIGER/safety-track). In addition, this year for the first time there is a track based on full LTL specifications in TLSF (in the following: TLSF/LTL-track). Each track is divided into subtracks for *realizability checking* and *synthesis*, and into two execution modes: *sequential* (using a single core of the CPU) and *parallel* (using up to 4 cores). We explain the rules of the competition, including evaluation of tools in the different tracks.

Submissions. Participants hand in their tools as source code, with installation instructions and a short description of the algorithms and optimizations used to solve the synthesis problem. Each author can submit up to three configurations of a tool per subtrack and execution mode. Tools are tested on a subset of the benchmarks before the competition, and we allow bugfixes in case of crashes or wrong results, if time permits.

Ranking Scheme. In all tracks, a correct answer within the timeout of 3600s is rewarded with one point for the solver, and a wrong answer is punished by subtracting 4 points. In the realizability tracks, correctness is determined by the realizability information stored in the files, if they have been used in previous competitions, or on a majority vote of the tools that solve the benchmark, otherwise. In the synthesis tracks, if the specification is realizable, then solution has to be model checked. This differs based on the input format, as explained below.

3.2 Specific Rules for TLSF/LTL-Track

Input Format. In the TLSF/LTL-track, specifications are given in basic TLSF [29]. The organizers supply the *synthesis format conversion* (SyFCo) tool⁶ that can be used to translate the specification to different existing specification formats. Specifications are interpreted according to standard LTL semantics, with respect to realizability as a Mealy machine.

Correctness of Solutions. In the synthesis subtrack, tools produce a solution in AIGER format if the specification is realizable. For syntactical correctness, the sets of inputs and outputs of the specification must be identical to the sets of inputs and outputs of the solution. Additionally, solutions are model checked against the specification with existing model checking tools. Only a solution that can be verified is counted as correct.

Legacy Tools. For comparison, we run legacy tool UNBEAST, non-competitive, in the realizability subtrack. To this end, the TLSF specification is translated to the native input format of UNBEAST, and a wrapper script converts outputs of the tool to the necessary format.

3.3 Specific Rules for AIGER/safety-Track

Input Format. In the AIGER/safety-track, specifications are given in the Extended AIGER Format for Synthesis [25, 27], modeling a single safety property.

Correctness of Solutions. In the synthesis subtrack, tools must produce a solution in AIGER format if the specification is realizable. For syntactical correctness, this solution must include the specification circuit, and must define how those inputs that are declared as `controllable` are computed (for details, see the SYNTCOMP 2014 report [27]). Additionally, these solutions are also model checked, and only verified solutions are counted as correct. To facilitate verification, synthesis tools can optionally output an inductive invariant that witnesses the correctness of the solution. If an invariant is supplied, we first try to determine correctness with an invariant check, and fall back to full model checking if the check is inconclusive.

Legacy Tools. For comparison, we run some of the entrants of SYNTCOMP 2014 and SYNTCOMP 2015 in the AIGER safety track. This allows us to highlight the progress of tools over the course of the last two years.

3.4 Selection of Benchmarks

AIGER/safety-track. In the AIGER-based track, the selection of benchmarks is based on information about the realizability and difficulty of benchmark problems that has been obtained from the results of previous competitions. This information is stored inside the benchmark files, as described in the SYNTCOMP 2015 report [28]. From each class of benchmarks, we selected a number of problems with an even distribution over difficulties (in terms of the ratio of solvers from previous competitions that were able to solve the benchmark). For benchmarks that have been added in 2016, we estimated their difficulty and realizability based on preliminary experiments with solvers from SYNTCOMP 2015.

⁶SyFCo is available at <https://github.com/reactive-systems/syfco>. Accessed August 2016.

The number of selected problems from each category (cp. Section 2) is given in Table 1. Compared to SYNTCOMP 2015, we

- replaced the existing HWMCC benchmarks completely with a selection from the new HWMCC set, since the existing ones were very hard and most instances could not be solved at all,
- replaced the existing $1t12dba/1t12dpa$ (LTL2AIG) benchmarks with the new LTL2DBA/LTL2DPA benchmarks, since the existing instances were very easy and in 2015, all participants solved all instances,
- reduced the number of benchmarks from the toy examples from 8 to 5 per example, and
- changed the selection for some existing benchmark classes to make it more challenging.

Table 1: Number of selected Benchmarks per Category, AIGER/safety-track

Category	Benchmarks	Category	Benchmarks
AMBA	16	Genbuf (LTL2AIG)	8
(Washing) Cycle Scheduling	16	Add (Toy Examples)	5
Demo (LTL2AIG)	16	Count (Toy Examples)	5
Driver Synthesis	16	Bitshift (Toy Examples)	5
Factory Assembly Line	15	Mult (Toy Examples)	5
Genbuf	16	Mv/Mvs (Toy Examples)	5
HWMCC	16	Stay (Toy Examples)	5
HyperLTL	16	Huffman Encoder	5
Load Balancer (LTL2AIG)	16		
LTL2DBA/LTL2DPA	16		
Moving Obstacle	16		
Matrix Multiplication	16	Total:	234

Like last year, in the synthesis subtracks we only considered benchmark instances that have been solved by at least one participant in the realizability track, resulting in a selection of 215 of the instances above.

TLSE/LTL-track. In the TLSE-based track, for realizability checking we have used all of the non-parameterized benchmarks, and have scaled up the parameter of the parameterized benchmarks until none of the tools was able to solve them (or we deemed it very likely that the next value could be solved by any of the tools).⁷ Overall, 195 instances were used for realizability checking. For synthesis, we excluded all the benchmarks for which none of the competitors could determine realizability, resulting in an overall set of 185 instances.

3.5 Execution

Like in the previous year, SYNTCOMP 2016 was run at Saarland University, on a set of identical machines with a single quad-core intel Xeon processor (E3-1271 v3, 3.6GHz) and 32 GB RAM (PC1600,

⁷Note that this makes comparison to tools that did not participate in the competition slightly non-trivial: if the additional tool can solve one of the parameterized benchmarks for higher parameters, then an approximately correct solution is to count these additional solutions towards the additional tool, while assuming that none of the other tools can solve them.

ECC), running a GNU/Linux system. Each node has a local 480 GB SSD that can be used as temporary storage.

Also like in previous years, the competition was organized on the EDACC platform [4], with a very similar setup. To ensure a high comparability and reproducibility of our results, a complete machine was reserved for each job, i.e., one synthesis tool (configuration) running one benchmark. Olivier Roussel’s *runsolver* [41] was used to run each job and to measure CPU time and wall time, as well as enforcing timeouts. As all nodes are identical and no other tasks were run in parallel, no other limits than a timeout of 3600 seconds (CPU time in sequential mode, wall time in parallel mode) per benchmark was set. Like last year, we used wrapper scripts to execute solvers that did not conform completely with the output format specified by the competition, e.g., to filter extra information that was displayed in addition to the specified output.

The model checker used for checking correctness of solutions for the AIGER/Safety track is IIMC⁸ in version 2.0. This year, we also allowed the safety solvers to additionally output a winning region of the safety game, which should be an inductive invariant for the synthesized circuit. If this information was supplied, then we first used an invariant check to determine correctness of the solution, and used full model-checking as a fallback solution if the invariant check failed.

For the TLSF/LTL track, the model checker used was V3⁹ [48].

4 Participants

Overall, nine tools were entered into SYNTCOMP 2016: six in the AIGER/safety-track, and three in the TLSF/LTL-track. We briefly describe the participants and give pointers to additional information.

4.1 AIGER/safety-Track

This track had six participants in 2016: AbsSynthe, Demiurge and Simple BDD Solver were re-entered, and we received submissions of new tools SafetySynth, SDF and TermiteSAT. Participants AbsSynthe, Simple BDD Solver, SafetySynth and SDF are based on the classical backward-fixpoint-based approach to solving safety games using BDDs, and mainly differ in the implemented optimizations and heuristics. In contrast, Demiurge and TermiteSAT implement novel SAT-based approaches, inspired by corresponding approaches for finite-state model checking. An overview of the optimizations implemented in the different BDD-based tools is given in Table 2. For detailed explanations of the different optimizations, we refer to the SYNTCOMP 2014 report [27].¹⁰

Swiss AbsSynthe v2.0

The new version of AbsSynthe was submitted by R. Brenguier, G. A. Pérez, J.-F. Raskin, and O. Sankur, and competed in both the realizability and the synthesis track.

AbsSynthe implements different BDD-based synthesis approaches, combining the standard fixpoint computation for safety games with some or all of the following:

⁸IIMC is available at <ftp://vlsci.colorado.edu/pub/iimc/iimc-2.0.tar.gz>. Accessed August 2016.

⁹V3 is available at <https://github.com/chengyinwu/V3>. Accessed August 2016.

¹⁰In particular, note that “partitioned transition relation” in our case only means that a separate transition function is used for every latch of the circuit, but does not include further partitioning or introduction of auxiliary variables for shared parts (cp. [14]).

Table 2: Optimizations implemented in BDD-based Tools.

Technique	AbsSynthe	SafetySynth	SDF	Simple BDD Solver
automatic reordering	x	x	x	x
eager dereferencing of BDDs		x	x	x
direct substitution	x	x	x	x
partitioned transition relation	x	x	x	x
simultaneous conjunction and abstraction	x	x	x	x
compositional synthesis	x			
abstraction-refinement	x			x
co-factor based strategy extraction	x	x	x	N/A
forward reachability analysis	x			N/A
ABC minimization		x		N/A
additional optimizations	x		x	x

- decomposition of the problem into independent sub-games, with different methods for merging the games after solving them (as described by Brenguier et al. [13] and in the SYNTCOMP 2015 report [28]),
- abstraction-refinement algorithms that solve the game on overapproximations of the possible behaviors; this year, the abstraction algorithm is replaced with the one used also in Simple BDD Solver (see the SYNTCOMP 2014 report [27]), and
- portfolio approaches that run different algorithms (or algorithms with different settings) in parallel.

In sequential mode, three different algorithms were entered into the competition:

- configuration seq1 uses a standard BDD-based fixpoint computation with several optimizations (see Table 2), but without compositionality or abstraction,
- configuration seq2 uses the new abstraction algorithm, but no compositionality, and
- configuration seq3 use a compositional algorithm, combined with an abstraction method that falls back to the concrete game if no solution is found when reaching a fixed threshold.

In parallel mode, also three different methods competed:

- configuration par1 runs the three sequential configurations in parallel, plus one additional configuration that uses abstraction with fixed threshold, but no compositionality,
- configuration par2 runs four copies of seq1, only modified in the BDD reordering technique that is used (SIFT, WINDOW2, WINDOW3, or WINDOW4), and
- configuration par3 runs four copies of seq2, with the same set of different reordering techniques.

Strategy extraction in AbsSynthe uses the co-factor-based approach of Bloem et al. [7], with some additional optimizations as described in the SYNTCOMP 2014 report [27]. To facilitate verification of solutions, AbsSynthe also produces the winning region and includes it in the solution file. The algorithms used in AbsSynthe have been described in more detail by Brenguier et al. [12, 13].

Implementation, Availability. AbsSynthe is implemented in C++, and depends on the AIGER toolset¹¹ and the CUDD package for BDD manipulation (v2.5.1) [45].

The code is available at <https://github.com/gaperez64/AbsSynthe/tree/native-dev-par>.

Demiurge 1.2.0

Demiurge was submitted by R. Könighofer and M. Seidl, and competed in both the realizability and the synthesis track. Demiurge implements different SAT-based methods for solving the reactive synthesis problem. In the competition, three methods are used. One approximates the winning region of the system player by repeated SAT-calls that ask for states that can enter the error states in a single step, followed by a generalization step to find additional, similar states. The second is a re-implementation of the incremental induction approach to reactive synthesis, as described by Morgenstern et al. [37]. The third method encodes the reactive synthesis problem directly into a SAT problem, essentially supplying a template that fits all possible solutions, and asking the SAT solver to come up with the variable valuations that represent a winning strategy. On its own, none of the strategies can compete with the BDD-based methods. However, a combination of all three methods, where the current approximation of the winning region by the first two is used to restrict the solution template in the third approach, solved more problems than any other tool in the synthesis track of SYNTCOMP 2015 [28].

This year, Demiurge competed in exactly the same version as last year. For a more information on Demiurge, we refer to the original descriptions of the implemented algorithms [6, 44], as well as the SYNTCOMP 2015 report [28].

The code is available at http://www.iaik.tugraz.at/content/research/design_verification/demiurge/ under the GNU Lesser General Public License version 3.

Simple BDD Solver

Simple BDD Solver was submitted by L. Ryzhyk and A. Walker, and competed in the realizability track. Simple BDD Solver implements the standard BDD-based algorithm for safety games, including a large number of optimizations. In particular, it includes an abstraction-refinement approach inspired by de Alfaro and Roy [1], and is a simplified version of the solver that was developed for the Termite project¹² [43], adapted to safety games given in the AIGER format.

It uses the BDDs with dynamic variable reordering using the sifting algorithm [42], and a number of additional optimizations (again, cp. Table 2 and the SYNTCOMP 2014 report [27]). Three configurations entered the competition:

- a basic algorithm without abstraction (called basic),
- an algorithm with abstraction based on overapproximation of the winning region (called abs1), and
- an algorithm with abstraction based on both over- and underapproximation of the winning region (called abs2).

The first two configurations are essentially identical to the configurations that entered SYNTCOMP 2015, and the third one is new.

¹¹The AIGER toolset is available at <http://fmv.jku.at/aiger/>. Accessed August 2016.

¹²For more information on Termite, consult <http://www.termite2.org>. Accessed August 2016.

Implementation, Availability. Simple BDD Solver is written in the Haskell functional programming language. It uses the CUDD package for BDD manipulation (updated to v3.0.0 this year) and the Attoparsec Haskell package for fast parsing.¹³ Altogether, the solver, AIGER parser, compiler and command line argument parser are just over 300 lines of code.

The code is available online at: <https://github.com/adamwalker/syntcomp>.

SafetySynth

SafetySynth was submitted by L. Tentrup, and competed in both the realizability and the synthesis track. SafetySynth is a re-implementation of Realizer [27, 28] that implements the standard BDD-based algorithm for safety games, using the optimizations that were most beneficial for BDD-based tools in SYNT-COMP 2014 and 2015 (see Table 2). Unlike Realizer, SafetySynth supports co-factor based strategy extraction, including .

It competed in two configurations, which differed only in the reordering heuristics for BDDs: the *basic* version uses the GROUP_SIFT heuristics [39], while the *alternative* version uses LAZY_SIFT. To facilitate verification of solutions, SafetySynth also produces the winning region and includes it in the solution file.

Implementation, Availability. SafetySynth is written in Swift. It uses the CUDD package (v3.0.0) for BDD manipulation, the AIGER toolset for parsing input files, and ABC for reducing the size of solutions.

The code is available online at: <https://www.react.uni-saarland.de/tools/safetysynth/>.

SDF

SDF was submitted by A. Khalimov, and competed in both the realizability and synthesis track. SDF implements the standard BDD-based fixpoint algorithm, with a number of known optimizations, as well as some new heuristics.

Out of the well-known optimizations, it implements automatic reordering, eager dereferencing, direct substitution, partitioned transition relation, simultaneous conjunction and abstraction. In synthesis, it uses the co-factor-based extraction of winning strategies.

Additionally, it implements the following new heuristics:

- additional caching of BDDs for nodes of the specification circuit,
- storing information about variable orders in the BDD, and limiting reordering after some time,
- after determining realizability, all BDDs except the non-deterministic strategy are removed, and the strategy is re-ordered once more, and
- caching and re-use of AIGER circuits for BDD nodes during construction of the solution.

Implementation, Availability. SDF is written in C++. It uses the CUDD package in version 3.0.0.

The code is available online at: <https://github.com/5nizza/sdf>.

¹³Attoparsec is available at <https://hackage.haskell.org/package/attoparsec>. Accessed August 2016.

TermiteSAT

TermiteSAT was submitted by A. Legg, N. Narodytska and L. Ryzhyk, and competed in the realizability track. TermiteSAT implements a novel SAT-based method for synthesis of safety specifications, as well as portfolio and hybrid modes that run the new algorithm alongside one of the algorithms of Simple BDD Solver.

The basic idea of the SAT-based algorithm is to explore a subset of the concrete runs of the safety game and prove that these runs can be generalized to a concrete winning strategy for one of the players. In contrast to other existing synthesis methods, it does not store or compute the set of all winning states. Instead, it computes an approximation of the winning states during a counter-example-guided backtracking search for candidate solutions to bounded safety games. Information about the unbounded game can be extracted in the form of Craig interpolants if a SAT query shows that a candidate strategy permits no counter example run in the bounded game. The algorithm is explained in more detail in [36].

For SYNTCOMP 2016, the algorithm has also been combined with the BDD-based algorithm with abstraction from Simple BDD Solver in two variants:

- a simple portfolio approach that runs both algorithms in parallel, and
- a *hybrid* approach that shares information between the solvers by forwarding information about states that have been determined to be winning or losing from the SAT algorithm to the BDD algorithm. The alternate direction is not implemented so that in cases where the BDD representation of the winning states explodes, the SAT-based approach will not suffer.

Implementation, Availability. TermiteSAT is written in Haskell. It uses Glucose as SAT solver¹⁴ [2], and PeRIPLO¹⁵ [40] to find interpolants. The portfolio and hybrid modes include Simple BDD Solver and all its dependencies (see above).

The code is available online at: <https://github.com/alexlegg/TermiteSAT>.

4.2 TLSF/LTL-Track

Acacia4Aiger

Acacia4Aiger was submitted by R. Brenguier, G. A. Pérez, J.-F. Raskin, and O. Sankur, and competed in both the realizability and the synthesis track. Acacia4Aiger is an extension of the reactive synthesis tool Acacia+¹⁶, which solves the reactive synthesis problem for LTL specifications by a reduction to safety games, which can then be solved efficiently by symbolic incremental algorithms based on an *antichain* representation of sets of states. Additionally, it uses a *compositional* approach to solve conjuncts of a specification separately. Acacia+ has been described in more detail by Bohy et al. [9].

For SYNTCOMP 2016, Acacia+ has been extended with

- a translator from TLSF to its native input format, including the separation of sub-formulas into “spec units” that allows for compositional solving. This translator has been integrated into the SyFCo tool.

¹⁴Glucose is available at <http://www.labri.fr/perso/lisimon/glucose/>. Accessed August 2016.

¹⁵PeRIPLO is available at <http://verify.inf.usi.ch/periplo>. Accessed August 2016.

¹⁶Acacia+ is available at <http://lit2.ulb.ac.be/acaciaplus/>. Accessed August 2016.

- an encoding of the Mealy machines constructed during synthesis into AIGER circuits that conform to the SYNTCOMP rules. This encoding uses the Speculoos tools¹⁷ to parse the native Mealy machine output and generate a transducer in AIGER format.

Implementation, Availability. Acacia4Aiger is written in Python and C. It uses LTL2BA¹⁸ [23] to convert LTL specifications into Büchi automata. It also includes the Speculoos and SyFco tools, as mentioned above.

The code is available online at: <https://github.com/gaperez64/acacia4aiger>.

BoSy

BoSy was submitted by L. Tentrup, and competed in both the realizability and the synthesis track. BoSy uses the *bounded synthesis* approach [22] that solves the LTL synthesis problem by first translating the specification into a universal co-Büchi automaton, and then encoding acceptance of a system (with bounded number of states, but unspecified behavior) into a constraint system. In contrast to the originally proposed encoding into satisfiability modulo theories (SMT), BoSy uses an encoding into quantified Boolean formulas (QBF), as described by Faymonville et al. [19]. One advantage of this encoding is that it allows to keep the input valuations symbolic, and to solve the constraints without explicitly enumerating all possibilities (as existing SMT solvers do). To detect unrealizability, the existence of a bounded strategy of the environment to falsify the specification is encoded in a similar way, and checked in parallel.

The resulting QBF formulas have the quantifier structure $\exists\forall\exists$. Before solving them, they are first simplified, using the QBF preprocessor bloqger¹⁹. In realizability mode, the simplified formula is then directly solved, using the QBF solver RAReQS [30]. In synthesis mode, BoSy uses RAReQS only to get a satisfying assignment for the outer existential quantifier, and reducing the query to a 2QBF formula with quantifier structure $\forall\exists$. This 2QBF formula is then solved by the certifying QBF solver QuAbs [46], and the returned certificate represents a solution to the synthesis problem. This solution is then converted into AIGER format, and further simplified using the ABC framework.

Two configurations of BoSy competed in SYNTCOMP 2016, differing in how the bound for the implementation size is increased after finding that no solution exists for the current bound: one configuration increases the bound linearly, the other exponentially. BoSy supports both Mealy and Moore semantics natively, as well as the extraction of a winning strategy for the environment in case of unrealizable specifications.

Implementation, Availability. BoSy is written in Swift. It uses LTL3BA²⁰ [3] to convert LTL specifications into Büchi automata. It uses bloqger, RAReQS²¹ and QuAbs²² to solve QBF constraints, as mentioned above. Finally, it uses SyFco to translate TFSF specifications to its native input format, and ABC to simplify solutions.

The code is available online at: <https://www.react.uni-saarland.de/tools/bosy/>.

¹⁷The Speculoos tools are available at <https://github.com/romainbrenquier/Speculoos>. Accessed August 2016.

¹⁸LTL2BA is available at <http://www.lsv.ens-cachan.fr/~gastin/ltl2ba>. Accessed August 2016.

¹⁹Bloqger is available at <http://fmv.jku.at/bloqger>. Accessed August 2016.

²⁰LTL3BA is available at <https://sourceforge.net/projects/ltl3ba/>. Accessed August 2016.

²¹RAReQS is available at <http://sat.inesc-id.pt/~mikolas/sw/areqs/>. Accessed August 2016.

²²QuAbs is available at <https://www.react.uni-saarland.de/tools/quabs/>. Accessed August 2016.

PARTY-Elli

PARTY-Elli was submitted by A. Khalimov, and competed in both the realizability and the synthesis track. PARTY-Elli also uses the *bounded synthesis* approach [22] for solving LTL synthesis problems. It uses an encoding into SMT formulas, as described in the original approach. To detect unrealizability, it also checks for the existence of a strategy for the environment to falsify the specification, but only for very simple solutions, i.e., stateless strategies. The tool and the algorithms it implements have been described in more detail by Khalimov et al. [34, 35].

To solve the synthesis problem more efficiently, PARTY-Elli uses the following optimizations of the original approach:

- it uses a heuristic *strengthening* of assume-guarantee specifications: if the specification has both liveness and safety assumptions, then it first tries to synthesize a solution where safety guarantees have to hold whenever the safety assumptions hold, regardless of the liveness assumptions; should no solution exist, then it falls back to the full formula;
- to avoid duplication of work when increasing the bound, the SMT solver is used in incremental mode: whenever the bound increases, the constraints that are specific to the previous bound are retracted, and constraints for the new bound are added,
- an optimization that analyses the strongly connected components (SCCs) of the automaton that is obtained from the specification, and simplifies the encoding if there are multiple SCCs.

When the SMT solver finds a satisfying assignment to the constraints, the solution is first encoded into Verilog, and then converted to AIGER and simplified using existing tools.

Implementation, Availability. PARTY-Elli is written in Python. It uses LTL3BA to convert LTL specifications into Büchi automata. It uses Z3²³ [38] for solving SMT constraints. For converting generated Verilog code into AIGER, it uses v12mv (contained in VIS²⁴ [10] and ABC. Finally, it uses SyFCo to translate TLSF specifications to its native input format.

The code is available online at: <https://github.com/5nizza/party-elli>.

5 Experimental Results

We present the results of SYNTCOMP 2016, separated into the AIGER/safety-track and the TLSF/LTL-track. Both main tracks are separated into realizability and synthesis subtracks, and parallel and sequential execution modes. Detailed results of the competition are also directly accessible via the web-frontend of our instance of the EDACC platform at <http://syntcomp.cs.uni-saarland.de>.

5.1 AIGER/safety-Track: Realizability

In the track for realizability checking of safety specifications in AIGER format, 6 tools competed on 234 benchmark instances, selected from the different benchmark categories as explained in Section 3.4. Overall, 17 different configurations entered this track, with 11 using sequential execution mode and 6 using parallel mode. In the following, we compare the results of these 17 configurations on the 234

²³Z3 is available at <https://github.com/Z3Prover/z3>. Accessed August 2016.

²⁴VIS is available at <http://vlsci.colorado.edu/~vis/>). Accessed August 2016.

benchmarks selected for SYNTCOMP 2016. To assess improvements compared to previous competitions, we also give results of 4 of the best-performing tool configurations from SYNTCOMP 2014 and 2015.

We first restrict the evaluation of results to purely sequential tools, then extend it to include also the parallel versions, and finally give a brief analysis of the results.

Sequential Mode. In sequential mode, AbsSynthe competed with three configurations (seq1, seq2 and seq3), Demiurge with one configuration (D1real), Simple BDD Solver with three configurations (basic, abs1, abs2), SafetySynth with two configurations (basic and alternative), as well as SDF and TermiteSAT with one configuration each.

The number of solved instances per configuration, as well as the number of uniquely solved instances, are given in Table 3. No tool could solve more than 175 out of the 234 instances, or about 75% of the benchmark set. 20 instances could not be solved by any tool within the timeout. For comparison, we also ran a number of additional tools on the benchmark set: the best-performing sequential configuration of AbsSynthe from SYNTCOMP 2015, as well as the two winning configurations from SYNTCOMP 2014 and 2015, both from Simple BDD Solver.²⁵ The results for these tools are grayed out in the table, and are not counted towards unique solutions.

Table 3: Results: AIGER Realizability (sequential mode only)

Tool	(configuration)	Solved	Unique
Simple BDD Solver	(abs1)	175	1
Simple BDD Solver	(abs2)	167	1
SafetySynth	(basic)	164	0
Simple BDD Solver	(basic)	164	0
SafetySynth	(alternative)	163	0
Simple BDD Solver	(2014)	163	
Simple BDD Solver	(2015, 2)	163	
AbsSynthe	(seq3)	159	4
AbsSynthe	(2015, seq2)	159	
AbsSynthe	(seq2)	149	0
SDF		149	0
AbsSynthe	(seq1)	145	0
Demiurge	(D1real)	129	6
TermiteSAT		97	4

Figure 1 gives a cactus plot for runtimes of the best-performing sequential configuration of each tool.

Parallel Mode. Three of the tools that entered the competition had at least one parallel configuration for the realizability track: three configurations of AbsSynthe (par1, par2, par3), one configuration of Demiurge (P3real), and two configurations of TermiteSAT (portfolio, hybrid). The parallel configurations had to solve the same set of benchmark instances as in the sequential mode. In contrast to the

²⁵We do not consider last year's version of Demiurge separately, since it participates in the same version as last year.

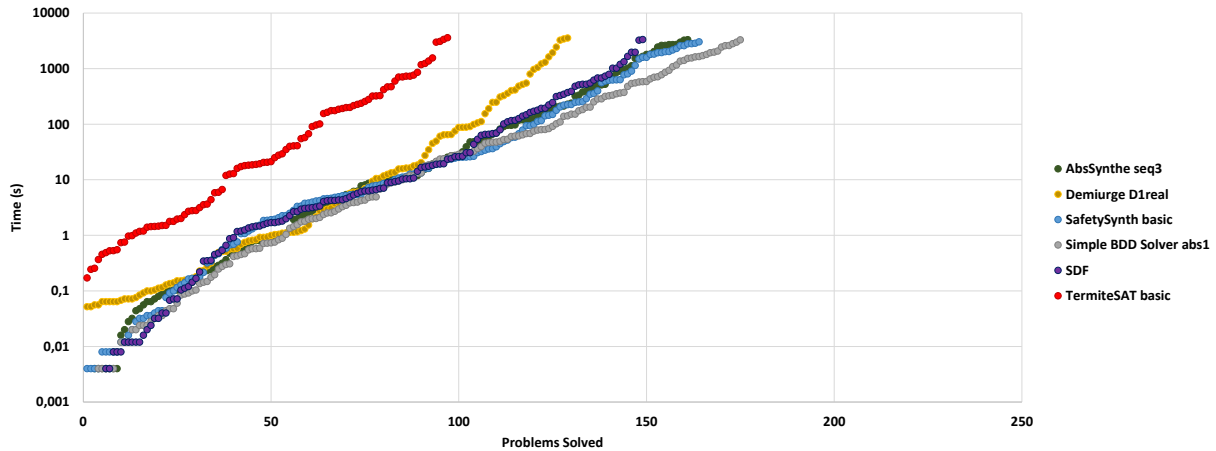


Figure 1: Runtime Cactus Plot of Best Sequential Configurations

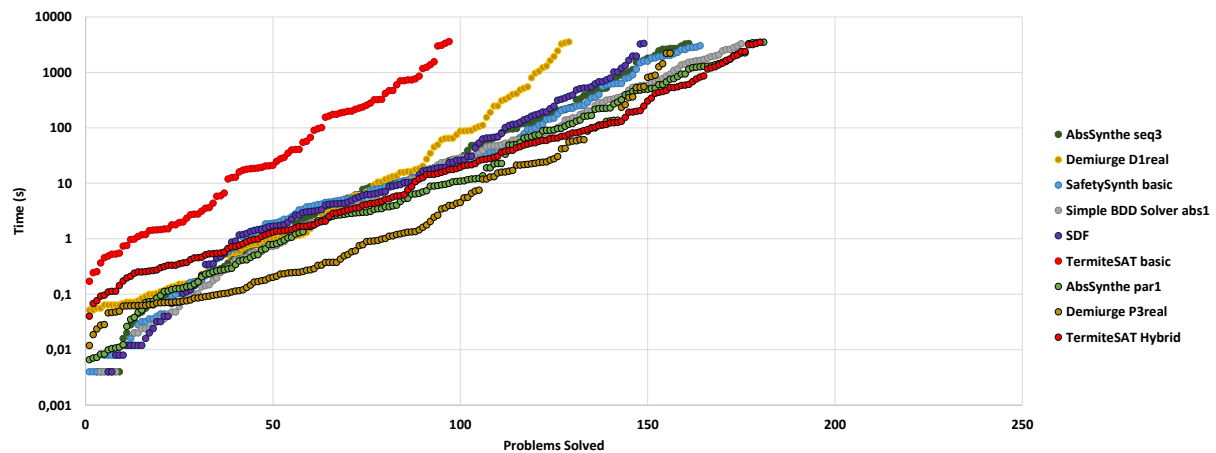


Figure 2: Runtime Cactus Plot of Best Configurations Overall

sequential mode, runtime of tools is now measured in wall time instead of CPU time. The results are given in Table 4. Compared to sequential mode, a number of additional instances could be solved: both AbsSynthe and TermiteSAT have one or more configurations that solve more than the best tool in sequential mode (about 77% of the benchmark set). Only 14 instances could not be solved by any tool in either sequential or parallel mode.

Note that in Table 4 we only count a benchmark instance as uniquely solved if it is not solved by any other configuration, including the sequential configurations. Still, AbsSynthe (par1) can provide one unique solution, and Demiurge (P3real) five.

We also ran this year's benchmark selection on the best parallel configuration from SYNTCOMP 2015. The results for AbsSynthe (2015, par1) appear grayed out in the table. This was the only parallel configuration that we ran from previous versions, since AbsSynthe was the only participating tool with a parallel mode that ran in SYNTCOMP 2015 and was not the same as one of the configurations that ran this year, and in SYNTCOMP 2014 the best configuration overall was a sequential configuration.

Table 4: Results: AIGER Realizability (parallel mode only)

Tool	(configuration)	Solved	Unique
AbsSynthe	(par1)	181	1
TermiteSAT	(hybrid)	180	0
TermiteSAT	(portfolio)	179	0
AbsSynthe	(2015, par1)	172	
Demiurge	(P3real)	156	5
AbsSynthe	(par3)	148	0
AbsSynthe	(par2)	141	0

Both modes: Solved Instances by Category. For both the sequential and the parallel configurations, Figures 3 and 4 give an overview of the number of solved instances per configuration and category (as defined in Table 1).

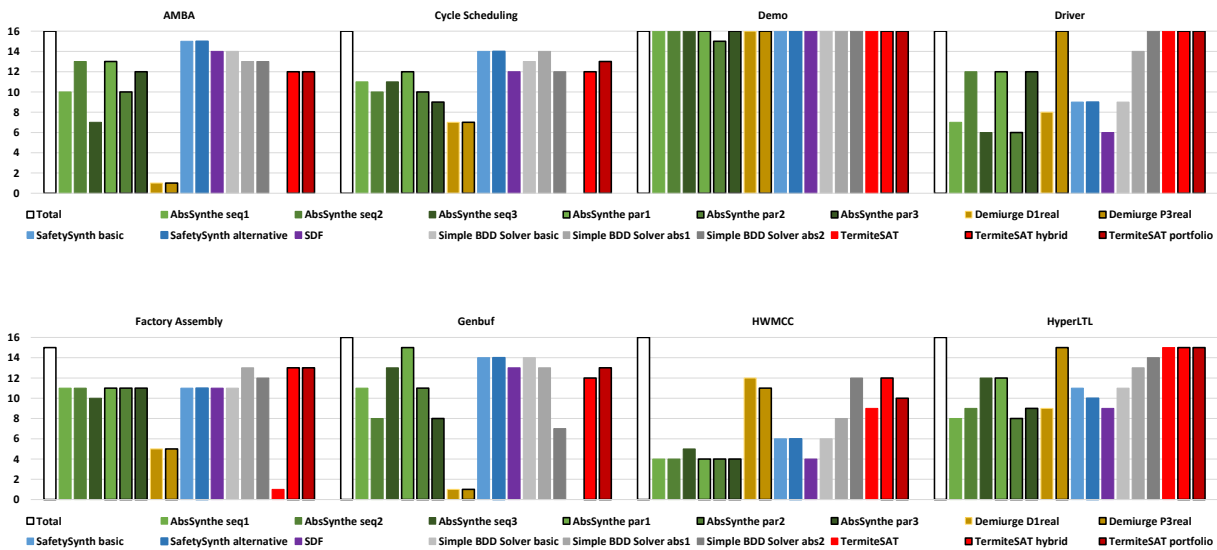


Figure 3: AIGER/safety Realizability Track, Solved Instances by Category (part 1)

Analysis. We note that on this year’s benchmark set, Simple BDD Solver (abs1) can solve significantly more problems than its best configuration from last year. Since these two configurations use the same algorithm, the main difference between them is the update to CUDD v3.0.0, which seems to give a significant performance boost. The new algorithm Simple BDD Solver (abs2) that uses also an underapproximation of the winning states cannot solve the same number of problems, but does solve a number of problems that Simple BDD Solver (abs1) does not solve, including one unique solution among all sequential configurations.

SafetySynth solves fewer problems than Simple BDD Solver (abs1 and abs2), but its basic configuration still solves one additional instance, compared to last year’s winner, Simple BDD Solver (2015, 2). The alternative version with a different BDD reordering scheme shows a very similar behavior, and

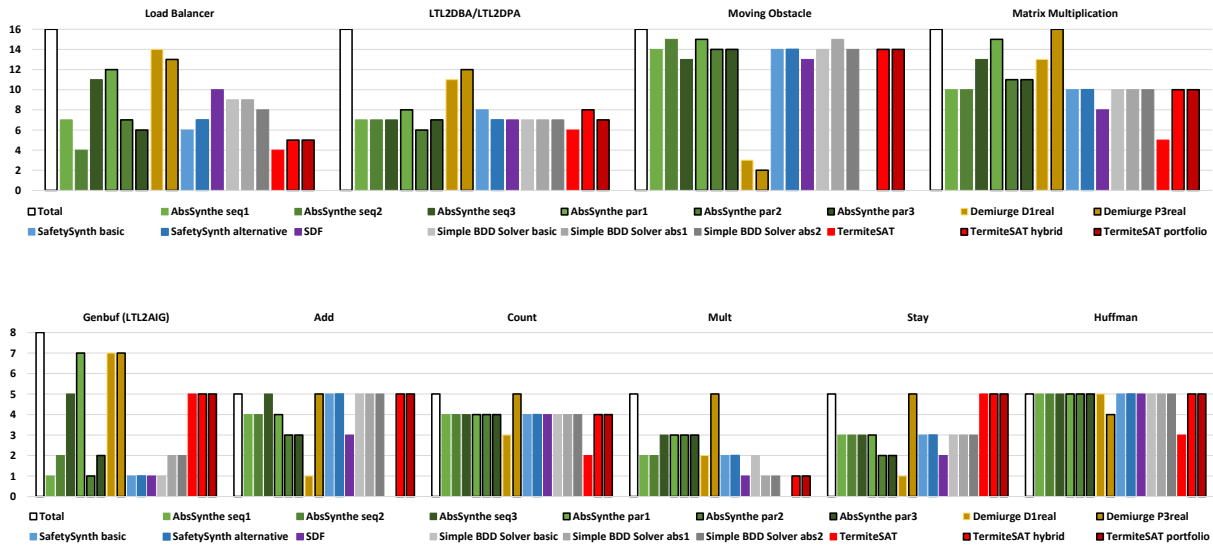


Figure 4: AIGER/safety Realizability Track, Solved Instances by Category (part 2)

solves one instance less overall. Note that SafetySynth is also based on CUDD v3.0.0, so it presumably benefits from the same performance boost as Simple BDD Solver, compared to tools that use an earlier version.

The best sequential configuration of AbsSynthe solves exactly the same number of problems as its best configuration last year, and less than both Simple BDD Solver and SafetySynth. Comparing the different sequential configurations, we note that they have similar behavior on many categories, but are also substantially different on some, e.g. AMBA, Driver, and Genbuf (LTL2AIG) (cp. Figure 4). Overall, the configuration seq3 that uses both compositionality and abstraction solves the highest number of problems, and the configuration seq1 that uses neither solves the least number. The parallel configuration AbsSynthe (par1) solves the highest number of problems among all configurations, and shows that a well-chosen portfolio approach can solve significantly more problems than any single algorithm. Note also that AbsSynthe still uses CUDD in version 2.5.1, so it could possibly benefit additionally from an upgrade to the newer version. The parallel configurations par2 and par3, which represent portfolios of seq1 and seq2, respectively, that only differ in the BDD reordering, solve fewer instances than their sequential counterparts.

We note that the new participant SDF solves the same number of problems as AbsSynthe (seq2). Finally, the SAT-based algorithms of Demiurge and TermiteSAT are in general not competitive in sequential mode, but each of them solves a significant number of problems uniquely. That is, they solve problems that none of the BDD-based algorithms can solve, but that are also not shared between the two SAT-based algorithms. In their parallel configurations, both tools show their strengths: TermiteSAT (hybrid) solves only one problem less than the best configuration overall, and Demiurge (P3real) solves a good number of instances, while supplying 5 solutions to problems that no other (sequential or parallel) configuration could solve. These additional problems can be attributed to the additional algorithms that are used in configuration P3real, including exchange of information between different algorithms. In contrast, the hybrid mode of TermiteSAT that also communicates information between algorithms did not produce unique solutions, and only solved one additional problem compared to the portfolio configuration without communication.

5.2 AIGER/safety-Track: Synthesis

In the track for synthesis from safety specifications in AIGER format, participants had to solve the same set of benchmarks as in the realizability track, except that we removed instances that could not be solved by at least one configuration in the realizability track. Thus, the benchmark set for the synthesis track contains 215 instances. Four tools entered the track: AbsSynthe with three sequential and three parallel configurations, Demiurge with one configuration for each mode, SafetySynth with two sequential configurations, and SDF with one sequential configuration.

For SYNTCOMP 2016, the ranking in the synthesis track is only based on the number of instances that can be solved within the timeout. The difference to the realizability track is that a solution for a realizable specification is only considered as correct if it can be model-checked within a separate timeout of one hour (cf. Section 3). As before, we start by presenting the results for the sequential configurations, followed by parallel configurations, and end with an analysis of the results.

Sequential Mode. In this mode, AbsSynthe competed with three configurations (seq1, seq2, seq3), Demiurge with one configuration (D1synt), SafetySynth with two configurations (basic, alternative), and SDF with one configuration.

Table 5 summarizes the experimental results, including the number of solved benchmarks, the uniquely solved instances, and the number of solutions that could not be model-checked within the timeout. Note that the “solved” column gives the number of problems that have either been correctly determined unrealizable, or for which the tool has presented a solution that could be verified. With this requirement, no sequential configuration could solve more than 153 or about 71% of the benchmarks, and 23 instances could not be solved by any tool. SDF is the only tool that generates a significant number of solutions that could not be verified.

Table 5: Results: AIGER Synthesis (sequential mode only)

Tool	(configuration)	Solved	Unique	MC Timeout
SafetySynth	(basic)	153	0	0
SafetySynth	(alt)	152	0	0
AbsSynthe	(seq3)	151	2	1
AbsSynthe	(seq2)	140	0	0
AbsSynthe	(seq1)	137	0	0
SDF		134	0	11
Demiurge	(D1synt)	118	2	3

Parallel Mode. In this mode, AbsSynthe competed with three configurations (par1, par2, par3), and Demiurge with one configuration (P3synt).

Table 6 summarizes the experimental results, again including the number of solved benchmarks, the uniquely solved instances, and the number of solutions that could not be verified within the timeout. No tool solved more than 165 problem instances, or about 77% of the benchmark set. There are only very few (potential) solutions that could not be verified within the timeout. Like in the parallel realizability track, we only consider instances as uniquely solved if they are not solved by any other configuration, including sequential ones. Still, Demiurge (P3Synt) was able to supply 9 unique solutions.

Table 6: Results: AIGER Synthesis (parallel mode only)

Tool	(configuration)	Solved	Unique	MC Timeout
AbsSynthe	(par1)	165	0	0
Demiurge	(P3Synt)	150	9	0
AbsSynthe	(par3)	140	0	1
AbsSynthe	(par3)	134	0	1

Analysis. Unsurprisingly, the number of solved instances for each tool in the synthesis track corresponds roughly to those solved in the realizability track. With respect to the smaller number of participants, both AbsSynthe (seq3) and Demiurge (D1synt) provide two unique solutions in the sequential mode, and Demiurge (D1synt) provides an additional 9 solutions that are unique over all sequential and parallel configurations.

As mentioned, SDF is the only tool that produces a significant number of solutions that cannot be verified. This is due to the fact that Demiurge usually produces small solutions, and both AbsSynthe and SafetySynth use the possibility to also produce a winning region that is used for verification. Thus, we can conclude that the introduction of additional witness information was successful and almost completely solves the problem of verification.²⁶

Even though for SYNTCOMP 2016 we do not have a ranking that is based on the size of solutions, we want to give a quick comparison. Figure 5 plots the sizes of synthesized strategies for some of the configurations. In contrast to previous size comparisons, we consider here not the size of the complete solution (which includes the specification circuit), but only the number of additional AND-gates, which corresponds to the strategy of the controller. Note that the solutions of Demiurge (P3synt) are often more than an order of magnitude smaller than the smallest BDD-based solutions, and there is sometimes another order of magnitude between the smallest and largest of those.

5.3 TLSF/LTL-Track: Realizability

In the track for realizability checking of LTL specifications in TLSF, 3 tools competed on 195 benchmark instances, selected as explained in Section 3.4. The three tools competed in 4 sequential and 2 parallel configurations. In the following, we compare the results of these 6 configurations on the 195 benchmarks selected for SYNTCOMP 2016. Additionally, we give a comparison to the legacy synthesis tool UNBEAST²⁷ [17], which was not entered as a participant, but only modified by the organizers in order to allow a rough comparison to the actual participants.

Again, we first restrict the evaluation of results to sequential configurations, then extend it to include parallel configurations, and finally give a brief analysis.

Sequential Mode. In sequential mode, BoSy competed with two configurations (lin and exp), and Acacia4Aiger and PARTY-Elli competed with one configuration each.

The number of solved instances per configuration, as well as the number of uniquely solved instances, are given in Table 7. No tool could solve more than 153 out of the 195 instances, or about 78% of

²⁶This is at least true in the current setup, where we first try a simple invariant check of the solution with respect to the winning region, and fall back to model checking if this check fails or times out.

²⁷UNBEAST is available at <http://www.react.uni-saarland.de/tools/unbeast/>. Accessed August 2016.

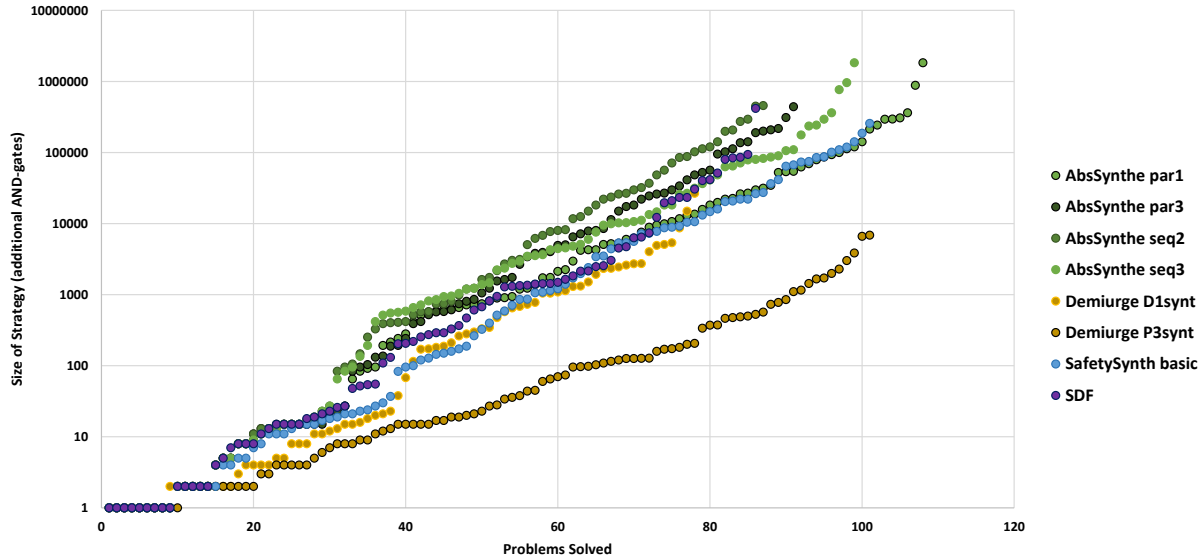


Figure 5: AIGER/safety Synthesis Track: Size of Solution Strategies for Selected Configurations

the benchmark set. 10 instances could not be solved by none of the participants within the timeout. Additional results for legacy tool UNBEAST are grayed out in the table, and are not counted towards solved instances or unique solutions.

Table 7: Results: TLSF Realizability (sequential mode only)

Tool	(configuration)	Solved	Unique
Acacia4Aiger		153	34
BoSy	(exp)	147	1
BoSy	(lin)	138	0
PARTY-Elli		118	0
UNBEAST		65	3

Figure 6 gives a cactus plot for runtimes of all sequential algorithms in the realizability track.

Parallel Mode. The only tool that also competed in parallel configurations was BoSy, with configurations lin par and exp par. As before, parallel configurations solve the same set of benchmark instances as in the sequential mode, but runtime is measured in wall time instead of CPU time. The results are given in Table 8. The parallel configurations of BoSy only solve a small number of additional instances.

Since in Table 8 we again only count a benchmark instance as uniquely solved if it is solved by neither another sequential nor another parallel configuration, there are no unique solutions.

Both modes: Solved Instances of Parameterized Benchmarks. For both the sequential and the parallel configurations, Figure 8 gives an overview of the number of solved instances per configuration and parameterized benchmark.

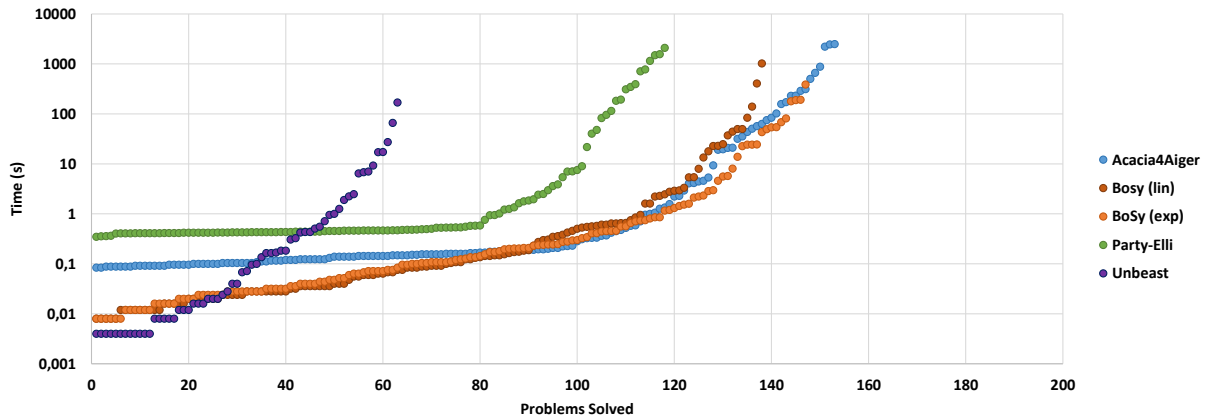


Figure 6: TLSF/LTL Realizability Track: Runtimes of Sequential Configurations

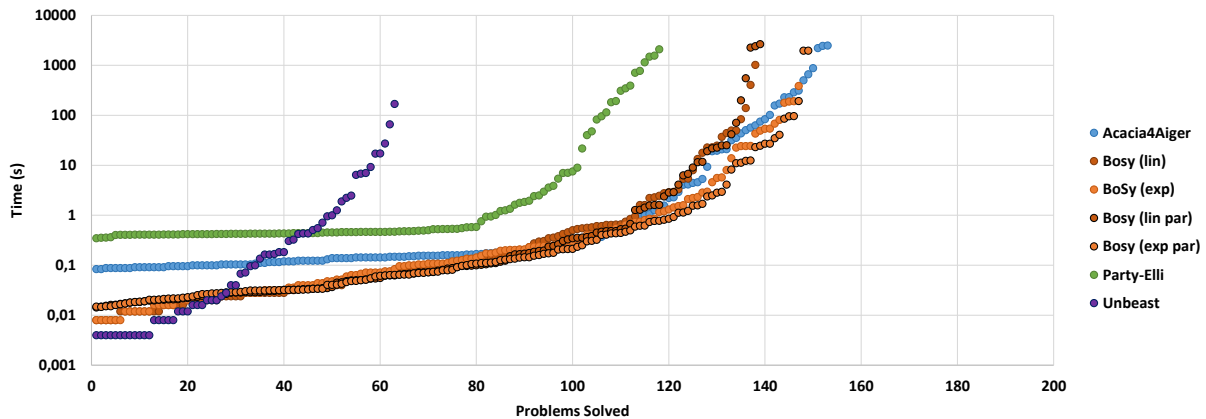


Figure 7: TLSF/LTL Realizability Track: Runtimes of All Configurations

Analysis. We note that Acacia4Aiger solves slightly more instances than BoSy (exp par), the most successful configuration that is based on a bounded synthesis algorithm. Since Acacia4Aiger has a high number of unique solutions, there must also be a high number of instances that is only solved by 2 or more configurations of the bounded synthesis tools. Thus, we conclude that the different algorithms have significantly different strengths. This also shows in the analysis of the parameterized benchmarks in Figure 8, where Acacia4Aiger clearly dominates for the benchmarks generalized buffer and the LTL2DBA benchmarks α , β , E, Q, and U1, while the bounded synthesis-based tools dominate on the round robin arbiter, the load balancer, and LTL2DBA benchmarks R and θ .

A remarkable result is that Acacia4Aiger has not solved a single unrealizable benchmark. This did not keep it from solving the highest number of instances, since only 20 of the 195 available benchmarks were unrealizable. Thus, a set of benchmarks that contained more unrealizable instances might have led to a very different result.

Also, PARTY-Elli did not solve a single instance of the load balancer benchmarks. A deeper analysis shows that this is due to the strategy applied in the strengthening optimization (see the description in Section 4.2): the strengthened formula is unrealizable, and the algorithm tries to find ever larger solutions for this unrealizable formula.

Table 8: Results: TLSF Realizability (parallel mode only)

Tool	(configuration)	Solved	Unique
BoSy	(exp par)	149	0
BoSy	(lin par)	139	0

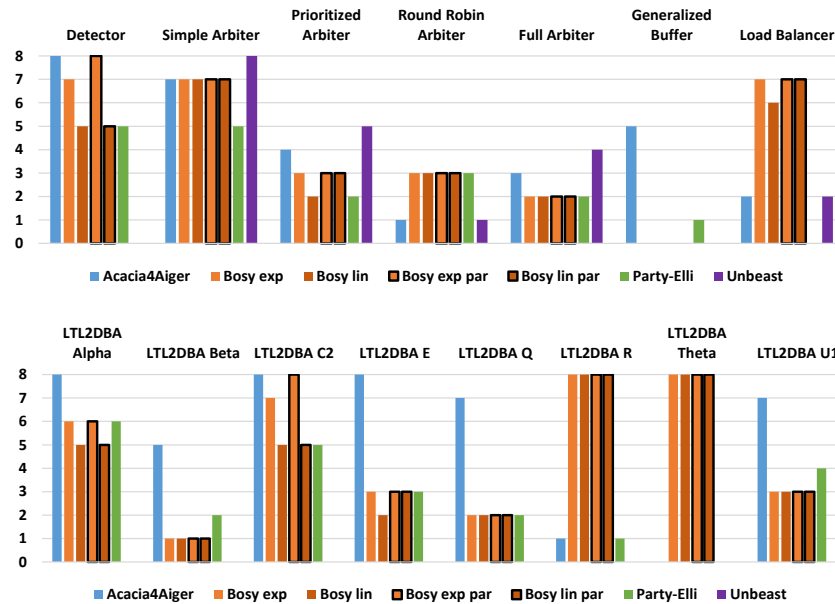


Figure 8: TLSF/LTL Realizability Track: Solved Instances for Parameterized Benchmarks

The parallel versions of BoSy were a bit more successful than the sequential versions, but the number of additionally solved instances is small.

Finally, legacy tool UNBEAST was not very successful overall, but did provide some unique solutions and solved the highest number of instances for some parameterized examples, namely the simple arbiter, prioritized arbiter, and full arbiter benchmarks.

5.4 TLSF/LTL-Track: Synthesis

In the track for synthesis from LTL specifications in TLSF, participants had to solve the same benchmarks as in the LTL/TLSF-realizability track, except that we removed instances that remained unsolved in the realizability track. Thus, the benchmark set for the synthesis track contains 185 instances. The track had the same participants as the LTL/TLSF-realizability track: Acacia4Aiger with one configuration, BoSy with four configurations, and PARTY-Elli with one configuration. Legacy tool UNBEAST did not run in this track, since it would have been an additional implementation effort to encode its native output into AIGER.

As for the AIGER/safety-track, the ranking is only based on the number of instances that can be solved within the timeout, and a solution for a realizable specification is only considered correct if it can be model-checked within a separate timeout of one hour (cf. Section 3). Again, we start by presenting the results for the sequential configurations, followed by parallel configurations, and end with an analysis

of the results.

Sequential Mode. Table 9 summarizes the experimental results for the sequential configurations, including the number of solved benchmarks, the uniquely solved instances, and the number of solutions that could not be model-checked within the timeout.

As before, the “solved” column gives the number of problems that have either been correctly determined unrealizable, or for which the tool has presented a solution that could be verified. With this requirement, no sequential configuration could solve more than 136 or about 74% of the benchmarks, and 22 instances could not be solved by any tool.

In this track, Acacia4Aiger is the only tool that generates solutions that could not be verified. Moreover, the submitted version of Acacia4Aiger produced 4 solutions that were determined to be incorrect by the model checker. The bug was fixed after the competition, and the results for the fixed version are also given in the table (grayed out). The 4 problems that are solved additionally by the fixed version are exactly those for which the originally submitted version produced a wrong solution.

Table 9: Results: TLSF Synthesis (sequential mode only)

Tool	(configuration)	Solved	Unique	MC Timeout	Wrong Solution
BoSy	(exp)	136	0	0	0
BoSy	(lin)	129	0	0	0
PARTY-Elli		119	6	0	0
Acacia4Aiger		133	16	20	4
Acacia4Aiger	(bugfix)	137		20	0

Parallel Mode. In this mode, BoSy competed with two configurations (lin par, exp par). Table 10 summarizes the experimental results, in the same format as before. No configuration solved more than 140 problem instances, or about 76% of the benchmark set. All solutions were verified within the timeout. As before, we only consider instances as uniquely solved if they are not solved by any other configuration, including sequential ones. Still, BoSy (exp par) provided 3 unique solutions.

Table 10: Results: TLSF Synthesis (parallel mode only)

Tool	(configuration)	Solved	Unique	MC Timeout	Wrong Solution
BoSy	(exp par)	140	3	0	0
BoSy	(lin par)	130	0	0	0

Analysis. As for the AIGER/safety-track, the number of solved instances for each tool in synthesis is closely related to the solved instances in realizability checking. The number of unique instances of Acacia4Aiger decreases significantly, in part due to the high number of solutions that could not be verified. As a consequence, we also note that PARTY-Elli and BoSy (exp par) now have a higher number of unique solutions. This suggests that these were problems that were solved by Acacia4Aiger and one

of the bounded synthesis tools in the realizability track, but in the synthesis track only the solution of the respective bounded synthesis tool could be verified.

Among the tools based on bounded synthesis, BoSy seems to benefit both from the QBF-based encoding (10 additional solutions when comparing BoSy (lin) against PARTY-Elli), and from the exponential search strategy (7, respectively 10, additional solutions when comparing the sequential, respectively parallel, versions with linear and exponential search strategy).

Even though for SYNTCOMP 2016 the tools are not ranked with respect to the size of solutions, we want to give a quick comparison. Figure 9 plots the sizes of solutions based on the number of AND-gates in the AIGER circuit, and Figure 10 plots the number of latches. Unsurprisingly, we see that solutions of bounded synthesis-based tools are usually smaller than those of Acacia4Aiger. However, there are also some differences in solution size between the solutions of BoSy and PARTY-Elli, pointing out that even if the synthesized solution is minimal with respect to some measure, there may still be differences in the concrete encoding to a circuit.

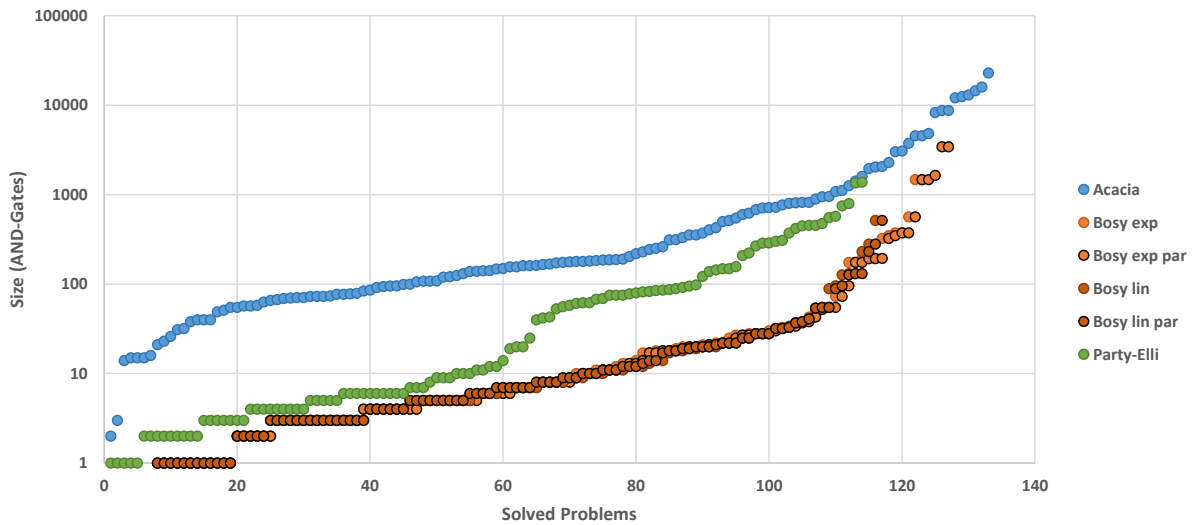


Figure 9: TLSF/LTL Synthesis Track: Solution Sizes, based on AND-Gates

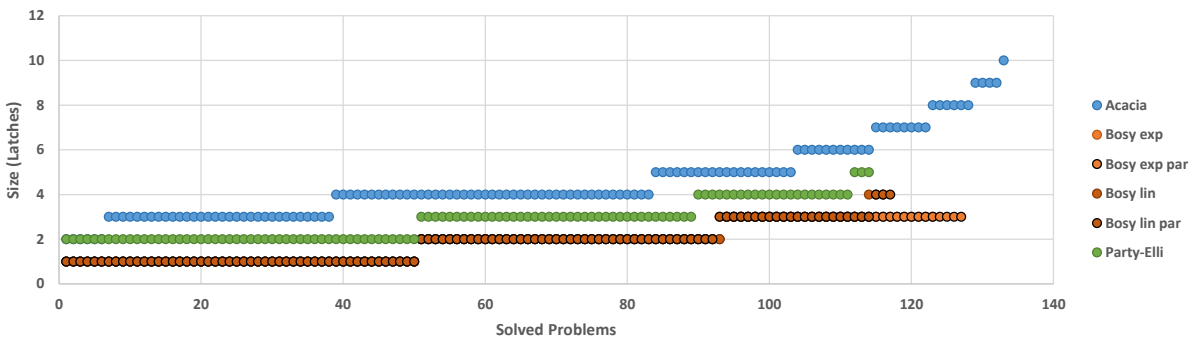


Figure 10: TLSF/LTL Synthesis Track: Solution Sizes, based on Latches

6 Conclusions

SYNTCOMP 2016 presented the first major extension of competition setup since the inception of the competition in 2014. In addition to a track based on specifications in the low-level AIGER format for safety properties, the third iteration of SYNTCOMP included for the first time a track based on specifications in the temporal logic synthesis format (TLSF) for full LTL properties.

In the pre-existing track, two new tools have been entered into the competition, and the results show significant improvements when compared to the best tools of last year. For the new track, three participants entered the competition that was executed for the first time with interesting results regarding the strengths and weaknesses of different existing approaches.

Acknowledgments. We thank Rüdiger Ehlers, Ioannis Filippidis, Andrey Kupriyanov, Kim Larsen, Nir Piterman, and Markus Rabe for interesting suggestions for the future of SYNTCOMP.

The organization of SYNTCOMP 2016 was supported by the Austrian Science Fund (FWF) through project RiSE (S11406-N23) and by the German Research Foundation (DFG) through project “Automatic Synthesis of Distributed and Parameterized Systems” (JA 2357/2-1), and its setup and execution by the European Research Council (ERC) Grant OSARES (No. 683300).

The development of AbsSynthe and Acacia4Aiger was supported by an F.R.S.-FNRS fellowship, and the ERC inVEST (279499) project.

The development of Demiurge was supported by the FWF through project RiSE (S11406-N23, S11408-N23).

The development of SafetySynth was supported by the ERC Grant OSARES (No. 683300).

The development of Simple BDD Solver was supported by a gift from the Intel Corporation.

NICTA is funded by the Australian Government through the Department of Communications and the Australian Research Council through the ICT Centre of Excellence Program.

References

- [1] Luca de Alfaro & Pritam Roy (2010): *Solving games via three-valued abstraction refinement*. *Inf. Comput.* 208(6), pp. 666–676, doi:10.1016/j.ic.2009.05.007.
- [2] Gilles Audemard & Laurent Simon (2009): *Predicting Learnt Clauses Quality in Modern SAT Solvers*. In: *IJCAI*, pp. 399–404. Available at <http://ijcai.org/Proceedings/09/Papers/074.pdf>.
- [3] Tomáš Babiak, Mojmir Kretínský, Vojtech Reháč & Jan Strejcek (2012): *LTL to Büchi Automata Translation: Fast and More Deterministic*. In: *TACAS, LNCS 7214*, Springer, pp. 95–109, doi:10.1007/978-3-642-28756-5_8.
- [4] Adrian Balint, Daniel Diepold, Daniel Gall, Simon Gerber, Gregor Kapler & Robert Retz (2011): *EDACC - An Advanced Platform for the Experiment Design, Administration and Analysis of Empirical Algorithms*. In: *LION 5. Selected Papers, LNCS 6683*, Springer, pp. 586–599, doi:10.1007/978-3-642-25566-3_46.
- [5] R. Bloem, B. Jobstmann, N. Piterman, A. Pnueli & Y. Sa’ar (2012): *Synthesis of Reactive(1) designs*. *J. Comput. Syst. Sci.* 78(3), pp. 911–938, doi:10.1016/j.jcss.2011.08.007.
- [6] R. Bloem, R. Könighofer & M. Seidl (2014): *SAT-Based Synthesis Methods for Safety Specs*. In: *VMCAI, LNCS 8318*, Springer, pp. 1–20, doi:10.1007/978-3-642-54013-4_1.
- [7] Roderick Bloem, Stefan J. Galler, Barbara Jobstmann, Nir Piterman, Amir Pnueli & Martin Weiglhofer (2007): *Specify, Compile, Run: Hardware from PSL*. *Electr. Notes Theor. Comput. Sci.* 190(4), pp. 3–16, doi:10.1016/j.entcs.2007.09.004.
- [8] Roderick Bloem, Swen Jacobs & Ayrat Khalimov (2014): *Parameterized Synthesis Case Study: AMBA AHB*. In: *SYNT, EPTCS 157*, pp. 68–83, doi:10.4204/EPTCS.157.9.
- [9] Aaron Bohy, Véronique Bruyère, Emmanuel Filiot, Naiyong Jin & Jean-François Raskin (2012): *Acacia+, a Tool for LTL Synthesis*. In: *CAV, LNCS 7358*, Springer, pp. 652–657, doi:10.1007/978-3-642-31424-7_45.

- [10] Robert K. Brayton, Gary D. Hachtel, Alberto L. Sangiovanni-Vincentelli, Fabio Somenzi, Adnan Aziz, Szu-Tsung Cheng, Stephen A. Edwards, Sunil P. Khatri, Yuji Kukimoto, Abelardo Pardo, Shaz Qadeer, Rajeev K. Ranjan, Shaker Sarwary, Thomas R. Shiple, Gitanjali Swamy & Tiziano Villa (1996): *VIS: A System for Verification and Synthesis*. In: *CAV, LNCS 1102*, Springer, pp. 428–432, doi:10.1007/3-540-61474-5_95.
- [11] Robert K. Brayton & Alan Mishchenko (2010): *ABC: An Academic Industrial-Strength Verification Tool*. In: *CAV, LNCS 6174*, Springer, pp. 24–40, doi:10.1007/978-3-642-14295-6_5.
- [12] Romain Brenguier, Guillermo A. Pérez, Jean-François Raskin & Ocan Sankur (2014): *AbsSynthe: abstract synthesis from succinct safety specifications*. In: *SYNT, EPTCS 157*, Open Publishing Association, pp. 100–116, doi:10.4204/EPTCS.157.11.
- [13] Romain Brenguier, Guillermo A. Pérez, Jean-François Raskin & Ocan Sankur (2015): *Compositional Algorithms for Succinct Safety Games*. In: *SYNT, EPTCS 202*, Open Publishing Association, pp. 98–111, doi:10.4204/EPTCS.202.7.
- [14] Jerry R. Burch, Edmund M. Clarke & David E. Long (1991): *Symbolic Model Checking with Partitioned Transition Relations*. In: *VLSI*, pp. 49–58.
- [15] Gianpiero Cabodi, Carmelo Loiacono, Marco Palena, Paolo Pasini, Denis Patti, Stefano Quer, Danilo Vendramineto, Armin Biere & Keijo Heljanko (2016): *Hardware Model Checking Competition 2014: An Analysis and Comparison of Solvers and Benchmarks*. *Journal on Satisfiability, Boolean Modeling and Computation* 9, pp. 135–172. Available at <https://www.satassociation.org/jsat/index.php/jsat/article/download/129/111>.
- [16] Alonzo Church (1962): *Logic, arithmetic and automata*. In: *Proceedings of the international congress of mathematicians*, pp. 23–35.
- [17] Rüdiger Ehlers (2011): *Unbeast: Symbolic Bounded Synthesis*. In: *TACAS, LNCS 6605*, Springer, pp. 272–275, doi:10.1007/978-3-642-19835-9_25.
- [18] Rüdiger Ehlers (2012): *Symbolic bounded synthesis*. *Formal Methods in System Design* 40(2), pp. 232–262, doi:10.1007/s10703-011-0137-x.
- [19] Peter Faymonville, Bernd Finkbeiner, Markus N. Rabe & Leander Tentrup (2015): *Encodings of Reactive Synthesis*. In: *Proceedings of QUANTIFY*. Available at <http://fmv.jku.at/quantify15/Faymonville-et-al-QUANTIFY2015.pdf>.
- [20] Bernd Finkbeiner & Swen Jacobs (2012): *Lazy Synthesis*. In: *VMCAI, LNCS 7148*, Springer, pp. 219–234, doi:10.1007/978-3-642-27940-9_15.
- [21] Bernd Finkbeiner, Markus N. Rabe & César Sánchez (2015): *Algorithms for Model Checking HyperLTL and HyperCTL**. In: *CAV, LNCS 9206*, Springer, pp. 30–48, doi:10.1007/978-3-319-21690-4_3.
- [22] Bernd Finkbeiner & Sven Schewe (2013): *Bounded synthesis*. *STTT* 15(5-6), pp. 519–539, doi:10.1007/s10009-012-0228-z.
- [23] Paul Gastin & Denis Oddoux (2001): *Fast LTL to Büchi automata translation*. In: *CAV, LNCS 2102*, Springer, pp. 53–65, doi:10.1007/3-540-44585-4_6.
- [24] Yashdeep Godhal, Krishnendu Chatterjee & Thomas A. Henzinger (2013): *Synthesis of AMBA AHB from formal specification: a case study*. *STTT* 15(5-6), pp. 585–601, doi:10.1007/s10009-011-0207-9.
- [25] Swen Jacobs (2014): *Extended AIGER Format for Synthesis*. *CoRR* abs/1405.5793. Available at <http://arxiv.org/abs/1405.5793>.
- [26] Swen Jacobs & Roderick Bloem (2016): *The Reactive Synthesis Competition: SYNTCOMP 2016 and Beyond*. In: *SYNT, Electronic Proceedings in Theoretical Computer Science 229*, Open Publishing Association, pp. 133–148, doi:10.4204/EPTCS.229.11.
- [27] Swen Jacobs, Roderick Bloem, Romain Brenguier, Rüdiger Ehlers, Timotheus Hell, Robert Könighofer, Guillermo A. Pérez, Jean-François Raskin, Leonid Ryzyk, Ocan Sankur, Martina Seidl, Leander Tentrup & Adam Walker (2016): *The First Reactive Synthesis Competition (SYNTCOMP 2014)*. *STTT*, doi:10.1007/s10009-016-0416-3. Published online first, journal issue to appear.

- [28] Swen Jacobs, Roderick Bloem, Romain Brenguier, Robert Könighofer, Guillermo A. Pérez, Jean-François Raskin, Leonid Ryzhyk, Ocan Sankur, Martina Seidl, Leander Tentrup & Adam Walker (2016): *The Second Reactive Synthesis Competition (SYNTCOMP 2015)*. In: *SYNT, EPTCS 202*, Open Publishing Association, pp. 27–57, doi:10.4204/EPTCS.202.4.
- [29] Swen Jacobs, Felix Klein & Sebastian Schirmer (2016): *A High-Level LTL Synthesis Format: TLSF v1.1*. In: *SYNT 2016, Electronic Proceedings in Theoretical Computer Science 229*, Open Publishing Association, pp. 112–132, doi:10.4204/EPTCS.229.10.
- [30] Mikolás Janota, William Klieber, Joao Marques-Silva & Edmund M. Clarke (2016): *Solving QBF with counterexample guided refinement*. *Artif. Intell.* 234, pp. 1–25, doi:10.1016/j.artint.2016.01.004.
- [31] Barbara Jobstmann (2007): *Applications and Optimizations for LTL Synthesis*. Ph.D. thesis, Graz University of Technology.
- [32] Barbara Jobstmann & Roderick Bloem (2006): *Optimizations for LTL Synthesis*. In: *FMCAD*, IEEE Computer Society, pp. 117–124, doi:10.1109/FMCAD.2006.22.
- [33] Ayrat Khalimov (2015): *Specification Format for Reactive Synthesis Problems*. In: *SYNT, EPTCS 202*, Open Publishing Association, pp. 112–119, doi:10.4204/EPTCS.202.8.
- [34] Ayrat Khalimov, Swen Jacobs & Roderick Bloem (2013): *PARTY Parameterized Synthesis of Token Rings*. In: *CAV, LNCS 8044*, Springer, pp. 928–933, doi:10.1007/978-3-642-39799-8_66.
- [35] Ayrat Khalimov, Swen Jacobs & Roderick Bloem (2013): *Towards Efficient Parameterized Synthesis*. In: *VMCAI, LNCS 7737*, Springer, pp. 108–127, doi:10.1007/978-3-642-35873-9_9.
- [36] Alexander Legg, Nina Narodytska & Leonid Ryzhyk (2016): *A SAT-Based Counterexample Guided Method for Unbounded Synthesis*. In: *CAV (2), LNCS 9780*, Springer, pp. 364–382, doi:10.1007/978-3-319-41540-6_20.
- [37] A. Morgenstern, M. Gesell & K. Schneider (2013): *Solving Games Using Incremental Induction*. In: *IFM'13, LNCS 7940*, Springer, pp. 177–191, doi:10.1007/978-3-642-38613-8_13.
- [38] Leonardo Mendonça de Moura & Nikolaj Bjørner (2008): *Z3: An Efficient SMT Solver*. In: *TACAS, LNCS 4963*, Springer, pp. 337–340, doi:10.1007/978-3-540-78800-3_24.
- [39] Shipra Panda & Fabio Somenzi (1995): *Who are the variables in your neighborhood*. In: *ICCAD*, IEEE Computer Society / ACM, pp. 74–77, doi:10.1109/ICCAD.1995.479994.
- [40] Simone Fulvio Rollini, Leonardo Alt, Grigory Fedyukovich, Antti Eero Johannes Hyvärinen & Natasha Sharygina (2013): *PeRIPLO: A Framework for Producing Effective Interpolants in SAT-Based Software Verification*. In: *LPAR, LNCS 8312*, Springer, pp. 683–693, doi:10.1007/978-3-642-45221-5_45.
- [41] Olivier Roussel (2011): *Controlling a Solver Execution with the runsolver Tool*. *JSAT* 7(4), pp. 139–144. Available at http://jsat.ewi.tudelft.nl/content/volume7/JSAT7_12_Roussel.pdf.
- [42] Richard Rudell (1993): *Dynamic variable ordering for ordered binary decision diagrams*. In: *ICCAD*, IEEE Computer Society, pp. 42–47, doi:10.1145/259794.259802.
- [43] Leonid Ryzhyk, Adam Walker, John Keys, Alexander Legg, Arun Raghunath, Michael Stumm & Mona Vij (2014): *User-Guided Device Driver Synthesis*. In: *OSDI*, USENIX Association, pp. 661–676. Available at <https://www.usenix.org/conference/osdi14/technical-sessions/presentation/ryzhyk>.
- [44] M. Seidl & R. Könighofer (2014): *Partial witnesses from preprocessed quantified Boolean formulas*. In: *DATE'14*, IEEE, pp. 1–6. Available at <http://dx.doi.org/10.7873/DATE2014.162>.
- [45] Fabio Somenzi (1999): *Binary Decision Diagrams*. In: *Calculational system design*, 173, IOS Press, p. 303.
- [46] Leander Tentrup (2016): *Solving QBF by Abstraction*. CoRR abs/1604.06752. Available at <http://arxiv.org/abs/1604.06752>.
- [47] Cong Tian, Jun Song, Zhenhua Duan & Zhao Duan (2015): *LtlNfBa: Making LTL Translation More Practical*. In: *SOFL+MSVL, LNCS 9559*, Springer, pp. 179–194, doi:10.1007/978-3-319-31220-0_13.

- [48] Cheng-Yin Wu, Chi-An Wu, Chien-Yu Lai & Chung-Yang R. Haung (2014): *A Counterexample-Guided Interpolant Generation Algorithm for SAT-Based Model Checking*. *IEEE Trans. on CAD of Integrated Circuits and Systems* 33(12), pp. 1846–1858, doi:10.1109/TCAD.2014.2363395.