mypyvy: A Research Platform for Verification of Transition Systems in First-Order Logic

James R. Wilcox, Yotam M. Y. Feldman, Oded Padon, and Sharon Shoham

Abstract. `mypyvy` is an open-source tool for specifying transition systems in first-order logic and reasoning about them. `mypyvy` is particularly suitable for analyzing and verifying distributed algorithms. `mypyvy` implements key functionalities needed for safety verification and provides flexible interfaces that makes it useful not only as a verification tool but also as a research platform for developing verification techniques, and in particular invariant inference algorithms. Moreover, the `mypyvy` input language is both simple and general, and the `mypyvy` repository includes several dozen benchmarks—transition systems that model a wide range of distributed and concurrent algorithms. `mypyvy` has supported several recent research efforts that benefited from its development framework and benchmark set.

1 Introduction

`mypyvy` is an open-source[1] research platform for automated reasoning about symbolic transition systems expressed in first-order logic. A chief design goal for `mypyvy` is to lower the barrier to entry for developing new techniques for solver-aided analysis and verification of transition systems. As a result, `mypyvy`'s modeling language is simple and close to the underlying logical foundation, and the tool is designed as a collection of reusable components, making it easy to experiment with new verification techniques.

The main application domain of `mypyvy` is verification of complex distributed algorithms. Following prior work [33,34], transition systems in `mypyvy` are expressed in uninterpreted first-order logic (i.e., without theories). Using uninterpreted first-order logic is motivated by the experience that solvers often struggle when theories (e.g., arithmetic, arrays, or algebraic data types) are combined with quantifiers. Quantifiers are essential for describing distributed algorithms (e.g., to state properties about all messages in the network), but theories can often be avoided, yielding improved automation.

`mypyvy` consists of a language for expressing transition systems directly as logical formulas but in a convenient manner (Sec. 2), a tool for reasoning about such systems, and a collection of benchmarks accumulated over the last few years (Sec. 2.1). Fig. 1 depicts `mypyvy`'s components, which are divided to solver-based queries (Sec. 3) and invariant inference algorithms (Sec. 4). Solver-based queries such as inductiveness checking and bounded model checking are answered by translating them into satisfiability checks that are sent to external first-order solvers. These queries are used as basic building blocks for developing invariant inference algorithms. `mypyvy` includes

[1] `https://github.com/wilcoxjay/mypyvy`

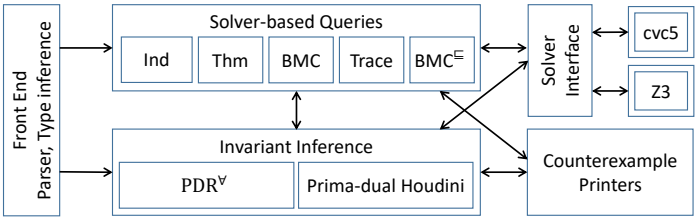

Fig. 1. Main components of `mypyvy`.

an implementation of two such algorithms: PDR$^\forall$ [21] and Primal-dual Houdini [35]. `mypyvy`'s internals are designed with the goal of making it easy to build on (Sec. 5). `mypyvy` interacts with multiple solvers, and currently supports Z3 [13] and cvc5 [2]. To present counterexamples (states, transitions, or traces) in a user-friendly way, `mypyvy` supports custom printers that simplify and improve readability of counterexamples.

`mypyvy` is not just the sum of the analyses currently available; it is a platform for doing research in automated verification. Several projects (including ongoing ones) use the `mypyvy` foundation and benchmark suite to build new invariant inference technique, user interfaces for verification and exploration, and, most recently, liveness verification techniques (Sec. 6).

`mypyvy`'s first-order modeling is inspired by Ivy [30,34], which promoted the idea of modeling distributed systems in the EPR decidable fragment of first-order logic. Ivy includes a rich and modular high-level imperative specification language, as well as mechanisms for creating executable implementations, specification-based testing, liveness verification, and more. As a result, Ivy's syntax, semantics, and code base are more complicated than what would be ideal for enabling rapid exploration of new techniques. In contrast, `mypyvy`'s focus on transition systems, with a simple syntax and semantics, makes it especially suited for enabling verification research.[2] Moreover, `mypyvy`'s code base is intentionally designed, documented, and typed (using Python's support for type annotations), to make it easy to build on and extend.

Broadly, `mypyvy` has three target audiences:

1. Researchers interested in modeling and verifying distributed algorithms. `mypyvy` offers a user-friendly input language, several queries that assist in developing models of distributed algorithms, readable counterexamples, and access to a variety of automatic verification algorithms.
2. Researchers developing verification techniques, and invariant inference in particular. `mypyvy` offers a starting point for implementing new algorithms on top of a developer-friendly code base. `mypyvy` includes many useful building blocks, and has already been successfully used in several research projects.
3. Researchers looking for benchmarks for various verification tasks. `mypyvy` includes a significant set of transition systems (and their invariants), which can serve as benchmarks for invariant inference or other verification tasks.

[2] There are current open-source efforts to automatically translate Ivy to `mypyvy` [9,37], which would allow Ivy users to benefit from `mypyvy`'s algorithms.

```
sort node
sort value
sort quorum

immutable relation member(node, quorum)
axiom forall Q1, Q2. exists N.
         member(N, Q1) & member(N, Q2)

mutable relation v(node, value)
mutable relation b(node)
mutable relation d(value)

init forall N, V. !v(N,V)
init forall N. !b(N)
init forall V. !d(V)

transition vote(n: node, x: value)
  modifies v, b
  !b(n) &
  (forall N, V.
    v'(N, V) <-> v(N, V) | (N = n & V = x)) &
  (forall N. b'(N) <-> b(N) | N = n)

transition decide(x: value)
  modifies d
  (exists Q. forall N. member(N, Q) -> v(N, x)) &
  (forall V. d'(V) <-> d(V) | V = x)

safety [agreement] forall X, Y. d(X) & d(Y) -> X = Y
invariant [decision_quorums] forall X. d(X) ->
  exists Q. forall N. member(N, Q) -> v(N, X)
invariant [unique_votes] forall N, X, Y. v(N, X) & v(N, Y) -> X = Y
invariant [voting_bit] forall N, X. v(N, X) -> b(N)

zerostate theorem forall Q. exists N. member(N, Q)
onestate theorem unique_votes & decision_qurums -> agreement
twostate theorem forall N, X. voting_bit & vote(N, X) -> voting_bit'
```

Fig. 2. The toy consensus example in `mypyvy`.

```
unsat trace {
  vote
  vote
  vote
  decide
  decide
  assert !safety
}

sat trace {
  any transition
  assert exists N, V. v(N,V)
  decide
  assert exists V. d(V)
}
```

```
> mypyvy verify consensus.pyv

checking init:
 implies invariant agreement..ok.
checking transition vote:
 preserves invariant agreement..ok.
checking transition decide:
 preserves invariant agreement..no!

counterexample:
  universes:
    sort node (1): node0
    sort quorum (1): quorum0
    sort value (2): value0 value1

  immutable:
    member(node0,quorum0)

  state 0:
    d(value1)
    v(node0,value0)

  state 1:
    d(value0)
    d(value1)
    v(node0,value0)

error consensus.pyv: invariant
agreement is not preserved by
transition decide
```

Fig. 3. A counterexample to induction (CTI) for the toy consensus protocol's safety property without additional invariants.

2 Modeling Language

We present `mypyvy` through a simple example of modeling and analyzing a toy consensus protocol.[3] To get started, the user first expresses a transition system in `mypyvy`'s input language, which is a convenient syntax for (many-sorted) uninterpreted first-order logic. A `mypyvy` model of the toy consensus protocol is shown in Fig. 2. In this protocol, each node *votes* for a single value, and once a majority or *quorum* of nodes vote for the same value a *decision* takes place. Because majorities intersect, the protocol ensures that at most one value is decided on. Modeling an algorithm or system of interest as a transition system in first-order logic may involve some abstraction, e.g., modeling majorities as abstract quorums such that every two quorums intersect [32].

States. The first step is to choose the types over which the transition system is defined. In the fashion of first-order logic, the basic types are *uninterpreted sorts* (`mypyvy` does not use SMT theories). In the example, we use the sorts node, value, and quorum to represent the nodes that participate in the distributed system, the values they choose from, and the sets of nodes that suffice for a decision (we abstract

[3] While not useful as a consensus protocol, this example does illustrate important aspects from proofs of complex, widely used consensus protocols like Paxos [25].

majorities following [4,33]). The state of the system is modeled by variables which can be *constants* (individuals), *relations*, or *functions*, whose domains are constructed from the aforementioned sorts. Each state variable is either `immutable`, which means it does not change throughout an execution of the system, or `mutable`, which means it may change with each transition. In the example, all state variables are relations. An immutable relation member denotes membership of a node in a quorum. The other relations are mutable: v records votes of nodes for values, b tracks which nodes already voted, and d records decisions.

Axioms. `mypyvy` allows the user to define a "background theory" over the immutable symbols, which restricts the state space, via `axiom` declarations. In the example, the property that any two quorums intersect (abstracting majorities) is expressed as an axiom for the member relation (line 6). Another common background theory that is useful when modeling distributed protocols in `mypyvy` is a total order, which can be used to abstract the natural numbers in first-order logic (e.g., to model rounds or indices).

Initial states. The initial states are defined as those that satisfy all `init` declarations. In the example, these declare that all mutable relations are initially empty (lines 13 to 15).

Transitions. The transitions of the system are expressed by `transition` declarations. The semantics is that each transition executes atomically and can modify the system's state. Transitions can have parameters, which are local variables that are assigned nondeterministically whenever the transition is executed. The example has two transitions: vote(n, x) and decide(x) (lines 17 to 27). An important design choice of `mypyvy` is that the user specifies transitions by explicitly writing logical formulas. Each transition is defined over two states: variables in the usual notation refer to the state *before* the transition is applied (*pre-state*), and primed variables refer to the state *after* the transition (*post-state*). Pre-conditions are encoded as conjuncts in the formula about the pre-state; for example, vote requires that the node has not already voted by specifying !b(n). Post-conditions are encoded as conjuncts about the post-state, relating it to the pre-state; for example, vote specifies that the relation b is updated to include exactly the same nodes as before in addition to n. Writing transitions directly through formulas offers great flexibility, but in order to write these formulas succinctly, a transition starts with a `modifies` clause that declares which mutable state variables are changed by it. For any mutable state component *not* in the modifies clause, `mypyvy` implicitly adds a conjunct encoding that the component does not change. Formally, the transition relation is the disjunction of the formulas from each of the transitions, where parameters are existentially quantified.

Safety. Finally, the user may specify safety properties using first-order formulas in `safety` declarations. The agreement safety property in the example (line 29) states that at most one value is decided. A safety property holds if it is satisfied by every state that is reachable from the an initial state via a sequence of transitions.

2.1 Benchmarks

The `mypyvy` repository includes over 30 transition systems collected over the years. Some of these were translated from Ivy, while others were directly modeled in `mypyvy`. The benchmarks model a variety of distributed and concurrent algorithms, including consensus algorithms, networking algorithms, and cache coherence protocols. The variety of benchmarks, which also vary in complexity, is useful for evaluating and experimenting with new verification techniques.

3 Satisfiability-Based Queries

Once a transition system is specified, `mypyvy` supports several satisfiability-based queries over it, which are directly translated to satisfiability checks and handed off to solvers (currently Z3 [13] and cvc5 [2] are supported). These queries are useful building blocks for developing more advanced solver-aided algorithms, and for users who are interested in analyzing specific systems (especially during the model development process). For most queries, `mypyvy` provides counterexamples based on satisfying models obtained from solvers. And while solvers are not guaranteed to terminate, `mypyvy` makes it easy to follow the EPR fragment restrictions, which ensures termination.

3.1 Queries

Inductiveness checking. `mypyvy` allows the user to add `invariant` declarations to prove safety by induction. These are first-order formulas, whose conjunction (together with the safety properties) forms a candidate inductive invariant. Fig. 2 lists three supporting invariants (lines 30 to 33). The most common query in `mypyvy` is to check if the candidate invariant is inductive. When translating an inductiveness check to the solver, `mypyvy` splits it into one solver query per (transition, invariant) pair. In our experience, splitting the disjunction outside the solver improves performance and reliability, and, best of all, improves transparency for the user when one of the cases is more problematic (e.g., takes a long time).

Theorems. In addition to invariants, which are meant to hold in all reachable states of the transition system, `mypyvy` supports checking `theorem` declarations, which specify first-order formulas that are expected to be valid modulo the background theory (i.e., axioms). `zerostate` theorems refer to immutable state variables only, `onestate` theorems may refer to the mutable state variables as well, and `twostate` theorems involve two states, similarly to transitions. In the toy consensus example, a `zerostate` theorem (line 35) is used to state that quorums cannot be empty (follows from the quorum intersection axiom); a `onestate` theorem (line 36) is used to state that, given the background theory, the `unique_votes` and `decision_quorums` invariants imply the `agreement` safety property; and a `twostate` theorem (line 37) is used to check that the `voting_bit` invariant is preserved by the `vote` transition.

Bounded model checking (BMC). It is often useful to explore (un)reachability of a safety violation via BMC. Given a transition system and a safety property, BMC asks, "Is there a counterexample trace with $\leq k$ transitions?" BMC is implemented in the usual way, by unrolling the transition relation.

Trace queries. Trace queries allow the user to explore the possible executions of the system in a more targeted way than BMC. This is useful both when the user in interested only in specific scenarios, and when BMC does not scale to sufficient depth. As an illustration, in a model of a distributed system with many protocol steps, BMC may only reasonably scale to a small depth, say 5 transitions, but many interesting behaviors of the system may not occur until at least 10 or 15 transitions. In Fig. 2, lines 38 to 45 show a query for the nonexistence of an execution trace that starts with three vote transitions, followed by two decide transitions, and then reaches a safety violation. `mypyvy` translates such a query to a first-order formula that is checked for unsatisfiability.

As a complement of trace queries that are expected to be unsatisfiable (specified by the `unsat` keyword), it is also useful to make `sat` trace queries that are expected to be satisfiable, demonstrating that some behaviors are indeed possible.[4] For example, lines 47 to 51 show a query expecting the existence of a trace that starts with any transition after which there exists a vote, followed by a decide transition after which there exists a decision. (That is possible when the number of nodes is 1.) Such satisfiable trace queries are especially useful for detecting *vacuity bugs*, where, due to a modeling error, some transitions mistakenly cannot execute, potentially making the system erroneously safe.

Relaxed Bounded Model Checking ($BMC^{\sqsubseteq}$). So far we discussed *concrete* traces. `mypyvy` can also search for *relaxed* counterexample traces of a bounded depth. A relaxed trace consists of a sequence of interleaved transitions and "relaxation steps", where some elements get deleted from the structure. As shown in [21], a relaxed counterexample trace that starts at an initial state and ends in a safety violation *proves* that there is no universally quantified inductive invariant that implies safety. This is the case in the toy consensus example—a relaxed counterexample trace found by `mypyvy` for this example is provided in Appendix A. The key to implementing relaxed BMC queries is encoding universe reduction between states. `mypyvy` does so by introducing a mutable unary relation active for each sort and using it as a guard in every quantifier, effectively restricting the universe in each state to the "active" part. Relaxation steps are then modeled by adding a relax transition where each active relation in the post-state is a subset of the corresponding one in the pre-state (expressed as a universally quantified formula); all other state variables are unmodified over the active part. Finally, a relaxed BMC query is encoded similarly to a BMC query (with the added relax transitions), except that, due to the use of different active universes, the axioms are asserted not only at the beginning of the

[4] `mypyvy` uses solver queries to generate executions of the transition system. A solver is needed due to `mypyvy`'s flexible and abstract modeling language. More imperative modeling languages, e.g. that of Ivy, admit execution/simulation without solvers, which can be useful for invariant inference as well [40,42]. Such simulation can also be implemented for a fragment of `mypyvy`'s language.

trace but also after every (relaxation) step, together with assertions requiring that the active universe contains the constants and is closed under functions.

3.2 Counterexamples

When a query fails (except for a `sat trace` query), it is because the formula sent to the solver was satisfiable. In such cases, `mypyvy` obtains a model from the solver and displays a *counterexample*—which can be a state, a transition, or a trace, depending on the failing query. For example, when inductiveness checking fails, it returns either a 1-state model demonstrating a violation of safety at an initial state, or a 2-state model demonstrating a counterexample to induction (CTI). As another example, when BMC finds an execution that violates safety, it returns a k-state model providing a counterexample trace. Fig. 3 shows a CTI (2-state model) for the toy consensus protocol when the invariants supporting the safety property are omitted. In general, `mypyvy` displays a k-state model by first listing the universe of each sort and the interpretations of the immutable symbols (member in our example). Then, for each of the k states, the interpretations of the mutable symbols in that state are printed. For relations, by default `mypyvy` only prints positive literals, i.e., the tuples that are in the relation.

Annotations, plugins, and custom printers. In some cases, the default counterexample printing of `mypyvy` is not as readable as it could be. For example, if one of the sorts in the transition system is totally ordered (using a binary relation and suitable axioms), it would make sense to name the elements of that sort according to the total order. To improve the readability of counterexamples, `mypyvy` supports custom formatting via *printer plugins* and *annotations*. Every declaration in `mypyvy` can be tagged with *annotations*, which have no inherent meaning, but can be detected by plugins, e.g., to cause things to be printed differently. For example, the declaration `sort round @printed_by(ordered_by_printer, le)` invokes the `ordered_by_printer` plugin and tells `mypyvy` that the sort `round` should be printed in the order given by the `le` relation. `mypyvy` provides several other custom printers, including one for printing sorts that represent sets of elements coming from another sort (illustrated in Appendix A Fig. 5). Users can also implement their own custom printing plugins in Python.

`mypyvy` also supports a handful of other annotations. `@no_print` instructs `mypyvy` not to print a sort, relation, constant, or function at all, which can be useful either because of a custom printer for another symbol, or temporarily because the model is large and the symbol is irrelevant to the current debugging session. `@no_minimize` is used to instruct `mypyvy`'s model minimizer not to minimize elements of a certain sort or relation. The annotation framework is extensible, and we expect more uses for it to come up.

3.3 Decidability and Finite Counterexamples via EPR

In general, `mypyvy` does not restrict the quantifier structure used in formulas, nor the signatures of state variables. As a result, the first-order formulas that encode different queries in `mypyvy` are not guaranteed to reside in any decidable fragment and solvers may diverge. However, a common practice when working with `mypyvy` is to use the effectively propositional (EPR) [38,36] fragment of first-order logic, which imposes

certain restrictions on functions and quantifier alternations. To encode a system in EPR (i.e., ensure that formulas generated for all queries are in EPR), the user can rely on recently developed methodologies [33,39]. For example, the toy consensus example of Fig. 2 is in EPR. Satisfiability of EPR is decidable, and reliably checked by solvers. EPR enjoys a small-model property, which implies queries have finite counterexamples (if any). Solver reliability and finite counterexamples are key enablers for more advanced algorithms (e.g., invariant inference) that make thousands of solver queries and employ model-based techniques. `mypyvy`'s language is close to the underlying logic used in queries, making it relatively easy to follow the EPR restrictions.

4 Invariant Inference

`mypyvy`'s design aims to make it easy to implement complex solver-aided analysis algorithms on top of the simpler queries. Two such algorithms, for automatically finding inductive invariants, are included in `mypyvy`: PDR$^\forall$ and Primal-dual Houdini.

Universal Property-Directed Reachability (PDR$^\forall$). `mypyvy` includes an implementation of PDR$^\forall$ [21], which infers universally quantified inductive invariants in first-order logic. Like IC3/PDR [7], PDR$^\forall$ constructs invariants incrementally by finding backwards reachable states and "blocking" them relative to a "frame". To block a state, PDR$^\forall$ computes a "forbidden sub-state" that rules out all states containing a certain pattern. If PDR$^\forall$ succeeds, it returns the inductive invariant in the form of a conjunction of universally quantified clauses. Otherwise, it either loops forever or returns a relaxed trace, proving that no universally quantified inductive invariant exists for the property. On the toy consensus example, PDR$^\forall$ returns a relaxed trace similar to the one obtained by BMC$^\sqsubseteq$. `mypyvy`'s implementation is the state-of-the-art implementation of PDR$^\forall$, and was used for comparison with PDR$^\forall$ in various papers [23,35,40]. The results demonstrate the success of `mypyvy`'s PDR$^\forall$ implementation in solving benchmarks that only require universally quantified invariants.

Primal-Dual Houdini. Primal-dual Houdini [35] is a recent invariant inference algorithm based on a formal duality between reachability in transition systems and a notion of incremental induction proofs. `mypyvy` includes an implementation of Primal-dual Houdini for universally quantified invariants. Primal-dual Houdini works best for transition systems where the inductive invariant can be constructed incrementally, adding one universally quantified clause at a time. Several complex distributed algorithms have this feature. In cases where the invariant cannot be constructed incrementally, Primal-dual Houdini can find a witness for that fact. See [35] for more details and an empirical evaluation. Primal-dual Houdini was prototyped using `mypyvy`'s infrastructure, and its development is an example of the usefulness of `mypyvy` for research in invariant inference.

5 Designing `mypyvy`'s Internals

We designed `mypyvy`'s internals with the goal of making it easy to build on. The most important aspects of the internals from the developer's perspective are (1) using

typed Python, (2) the design of the abstract syntax trees (ASTs), and (3) the interface to the underlying first-order solver. `mypyvy` is written in statically typed Python using the `mypy` type checker. Types not only help catch bugs, but also document the interfaces available to the developer. In our experience, types allow developers to get up to speed more quickly on the code base and facilitate communication.

The ASTs for representing logical formulas in `mypyvy` were designed to support symbolic manipulation, as is common in solver-aided algorithms. This led us to avoid any additional intermediate representations between the ASTs representing the user-level formulas and the ASTs representing the input to solvers. We also structured the ASTs so that it is easy to (re)compute any analysis performed. For example, instead of using a traditional (mutable, long-lived) symbol table to resolve names, `mypyvy` uses a purely functional context to track scopes during AST traversals. The context is thrown away and recomputed every time the AST is traversed. This makes it easy to traverse programmatically generated ASTs, without needing to update any symbol tables or other global data structures, and the extra run time overhead is negligible.

Developers who use `mypyvy` often want to make many queries to the underlying solvers (currently Z3 and cvc5). We expose two interfaces for this. First, many common primitives, such as those discussed in Sec. 3.1, are exposed as a library. Second, `mypyvy` has a lower-level solver interface, where developers can issue their own satisfiability queries, and also gain access to minimized models and minimized unsat cores. Furthermore, developers of sophisticated invariant inference algorithms may have many thousands of queries to run, so `mypyvy` supports running many solvers in parallel.

6 Works Using `mypyvy`

One of `mypyvy`'s goals is to serve the research community and enable research on verification, and invariant inference in particular. Indeed, in recent years several works have built on `mypyvy` or used it to various extents.

Phase-PDR$^\forall$ [14] is a user-guided invariant inference technique. The user provides a *phase structure* to convey temporal intuition, and suitable *phase invariants* are found using an adaptation of PDR$^\forall$. Phase-PDR$^\forall$ was developed on top of the `mypyvy` code base and `mypyvy`'s PDR$^\forall$ implementation, and its evaluation uses benchmarks available from `mypyvy` augmented with phase structures.

SWISS [18] is an invariant inference algorithm that finds quantified invariants, including quantifier alternations, using explicit search. While SWISS does not use the `mypyvy` code base (it is implemented in C++), it accepts `mypyvy`'s input files and its evaluation uses benchmarks available from `mypyvy`.

P-FOL-IC3 [23] is a variant of IC3/PDR that can find invariants with arbitrary quantification using *quantified separation* [22]. P-FOL-IC3 was implemented using `mypyvy`'s code, and also benefited from `mypyvy`'s benchmark set.

IC3PO [15,16] is an IC3/PDR variant that finds quantified invariants for protocols by analyzing finite instances. It does not use `mypyvy`'s code, but is evaluated on some of `mypyvy`'s benchmarks, manually translated to its input format.

LVR [41] develops a methodology for proving liveness properties. It uses `mypyvy` "twice": first, as a modeling language and a source of benchmarks, and second, as an invariant inference engine (using P-FOL-IC3) to find invariants that are required to support a liveness proof based on ranking functions.

7 Related Work

Several tools promote specification and verification of systems and algorithms using first-order logic, dating back to Abstract State Machines [6,17]. Alloy [20] is a relational modeling language and a tool that performs bounded verification, i.e., bounding the size of the universe of each sort. Alloy goes beyond first-order logic and has concepts such as transitive closure, but it shares `mypyvy`'s emphasis on using uninterpreted relations and quantifiers, rather than SMT theories. Electrum [8,29] is an extension of Alloy that was recently integrated into Alloy 6 [1]; it essentially turns Alloy into a modeling language for transition systems. When universe sizes are bounded, Electrum/Alloy 6 can use finite-state model checkers to verify safety as well as liveness properties.

Ivy [31,34] is a multi-modal verification tool that supports modeling using first-order logic and EPR as well as some decidable SMT theories, modular reasoning, extracting executable implementations, liveness verification, specification-based testing, and more. Unlike Alloy, Ivy is not restricted to bounded verification; instead, it relies on user-provided inductive invariants and restricts the quantifier-alternation structure of verification conditions to ensure decidability of unbounded verification queries.

Verification of transition systems is also the focus of the TLA$^+$ toolbox [26], where transition systems are expressed in a very rich logic (based on set theory). As a result, verification is restricted to model checking bounded instances [24,43] similar to Alloy, or manually writing detailed machine-checked proofs [10].

The IronFleet project [19] verifies distributed systems by formalizing transition systems and refinement in Dafny [27], a general-purpose deductive verification language. In IronFleet, transition systems are expressed using the rich Dafny type system, which is based on SMT combined with quantifiers. But as a result, queries to Z3, the underlying SMT solver, suffer from instability, especially when quantifiers—which are handled using triggers—are involved [28].

Compared to the aforementioned systems, `mypyvy` takes a similar approach to Ivy in using first-order logic without theories and aiming for unbounded verification, but unlike Ivy it focuses on automatically finding inductive invariants, and enabling research in that direction. We note that automated invariant inference depends on the reliability of invariant checking and related queries, which is absent from Dafny, TLA+, or Alloy (for the unbounded case), and obtained in `mypyvy` by using EPR in the style of [33].

Another related line of research is developing intermediate representation languages for invariant inference. VMT [11] is a format that extends SMT-LIB [3] to a transition system semantics. Constrained Horn Clauses (CHCs) [5,12] is another SMT-LIB extension that is similar to transition systems but more general (it captures, e.g., recursive programs). Both VMT and CHCs are typically used with rich SMT theories, whereas `mypyvy`'s logic is centered around uninterpreted first-order logic and quantifiers.

References

1. Alloy 6 announcement (2021), https://alloytools.org/alloy6.html, accessed 2023-02-03
2. Barbosa, H., Barrett, C.W., Brain, M., Kremer, G., Lachnitt, H., Mann, M., Mohamed, A., Mohamed, M., Niemetz, A., Nötzli, A., Ozdemir, A., Preiner, M., Reynolds, A., Sheng, Y., Tinelli, C., Zohar, Y.: cvc5: A versatile and industrial-strength SMT solver. In: Fisman, D., Rosu, G. (eds.) Tools and Algorithms for the Construction and Analysis of Systems - 28th International Conference, TACAS 2022, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2022, Munich, Germany, April 2-7, 2022, Proceedings, Part I. Lecture Notes in Computer Science, vol. 13243, pp. 415–442. Springer (2022). https://doi.org/10.1007/978-3-030-99524-9_24, https://doi.org/10.1007/978-3-030-99524-9_24
3. Barrett, C., Stump, A., Tinelli, C.: The SMT-LIB Standard: Version 2.0. In: Gupta, A., Kroening, D. (eds.) Proceedings of the 8th International Workshop on Satisfiability Modulo Theories (Edinburgh, UK) (2010)
4. Berkovits, I., Lazic, M., Losa, G., Padon, O., Shoham, S.: Verification of threshold-based distributed algorithms by decomposition to decidable logics. In: Dillig, I., Tasiran, S. (eds.) Computer Aided Verification - 31st International Conference, CAV 2019, New York City, NY, USA, July 15-18, 2019, Proceedings, Part II. Lecture Notes in Computer Science, vol. 11562, pp. 245–266. Springer (2019). https://doi.org/10.1007/978-3-030-25543-5_15, https://doi.org/10.1007/978-3-030-25543-5_15
5. Bjørner, N.S., Gurfinkel, A., McMillan, K.L., Rybalchenko, A.: Horn clause solvers for program verification. In: Beklemishev, L.D., Blass, A., Dershowitz, N., Finkbeiner, B., Schulte, W. (eds.) Fields of Logic and Computation II - Essays Dedicated to Yuri Gurevich on the Occasion of His 75th Birthday. Lecture Notes in Computer Science, vol. 9300, pp. 24–51. Springer (2015). https://doi.org/10.1007/978-3-319-23534-9_2, https://doi.org/10.1007/978-3-319-23534-9_2
6. Börger, E., Stärk, R.F.: Abstract State Machines. A Method for High-Level System Design and Analysis. Springer (2003), http://www.springer.com/computer/swe/book/978-3-540-00702-9
7. Bradley, A.R.: Sat-based model checking without unrolling. In: Proceedings of the 12th International Conference on Verification, Model Checking, and Abstract Interpretation (VMCAI). Lecture Notes in Computer Science, vol. 6538, pp. 70–87. Springer (2011)
8. Brunel, J., Chemouil, D., Cunha, A., Macedo, N.: The electrum analyzer: model checking relational first-order temporal specifications. In: Huchard, M., Kästner, C., Fraser, G. (eds.) Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering, ASE 2018, Montpellier, France, September 3-7, 2018. pp. 884–887. ACM (2018). https://doi.org/10.1145/3238147.3240475, https://doi.org/10.1145/3238147.3240475
9. Chajed, T.: Ivy to mypyvy translator (2023), https://github.com/tchajed/ivy-to-mypyvy
10. Chaudhuri, K., Doligez, D., Lamport, L., Merz, S.: The tla$^+$ proof system: Building a heterogeneous verification platform. In: Cavalcanti, A., Déharbe, D., Gaudel, M., Woodcock, J. (eds.) Proceedings of the 7th International Colloquium on Theoretical Aspects of Computing (ICTAC). Lecture Notes in Computer Science, vol. 6255, p. 44. Springer (2010)
11. Cimatti, A., Griggio, A., Tonetta, S.: The VMT-LIB language and tools. CoRR **abs/2109.12821** (2021), https://arxiv.org/abs/2109.12821
12. De Angelis, E., K., H.G.V.: CHC-COMP 2022: Competition report. In: Hamilton, G.W., Kahsai, T., Proietti, M. (eds.) Proceedings 9th Workshop on Horn Clauses

for Verification and Synthesis and 10th International Workshop on Verification and Program Transformation, HCVS/VPT@ETAPS 2022, and 10th International Workshop on Verification and Program TransformationMunich, Germany, 3rd April 2022. EPTCS, vol. 373, pp. 44–62 (2022). https://doi.org/10.4204/EPTCS.373.5, https://doi.org/10.4204/EPTCS.373.5

13. De Moura, L., Bjørner, N.: Z3: An efficient SMT solver. In: TACAS. pp. 337–340 (2008)
14. Feldman, Y.M.Y., Wilcox, J.R., Shoham, S., Sagiv, M.: Inferring inductive invariants from phase structures. In: Proceedings of the 31st International Conference on Computer Aided Verification (CAV). Lecture Notes in Computer Science, vol. 11562, pp. 405–425. Springer (2019)
15. Goel, A., Sakallah, K.A.: On symmetry and quantification: A new approach to verify distributed protocols. In: Dutle, A., Moscato, M.M., Titolo, L., Muñoz, C.A., Perez, I. (eds.) NASA Formal Methods - 13th International Symposium, NFM 2021, Virtual Event, May 24-28, 2021, Proceedings. Lecture Notes in Computer Science, vol. 12673, pp. 131–150. Springer (2021). https://doi.org/10.1007/978-3-030-76384-8_9, https://doi.org/10.1007/978-3-030-76384-8_9
16. Goel, A., Sakallah, K.A.: Towards an automatic proof of lamport's paxos. In: Formal Methods in Computer Aided Design, FMCAD 2021, New Haven, CT, USA, October 19-22, 2021. pp. 112–122. IEEE (2021). https://doi.org/10.34727/2021/isbn.978-3-85448-046-4_20, https://doi.org/10.34727/2021/isbn.978-3-85448-046-4_20
17. Gurevich, Y.: Evolving Algebras 1993: Lipari Guide, pp. 9–36. Oxford University Press, specification and vgalidation methods edn. (January 1995), https://arxiv.org/pdf/1808.06255.pdf
18. Hance, T., Heule, M., Martins, R., Parno, B.: Finding invariants of distributed systems: It's a small (enough) world after all. In: Mickens, J., Teixeira, R. (eds.) 18th USENIX Symposium on Networked Systems Design and Implementation, NSDI 2021, April 12-14, 2021. pp. 115–131. USENIX Association (2021), https://www.usenix.org/conference/nsdi21/presentation/hance
19. Hawblitzel, C., Howell, J., Kapritsos, M., Lorch, J.R., Parno, B., Roberts, M.L., Setty, S.T.V., Zill, B.: IronFleet: proving practical distributed systems correct. In: Proceedings of the 25th ACM Symposium on Operating Systems Principles (SOSP). pp. 1–17. Monterey, CA (Oct 2015)
20. Jackson, D.: Software Abstractions: Logic, Language, and Analysis. MIT Press (Feb 2012)
21. Karbyshev, A., Bjørner, N., Itzhaky, S., Rinetzky, N., Shoham, S.: Property-directed inference of universal invariants or proving their absence. J. ACM **64**(1), 7:1–7:33 (2017)
22. Koenig, J.R., Padon, O., Immerman, N., Aiken, A.: First-order quantified separators. In: Donaldson, A.F., Torlak, E. (eds.) Proceedings of the 41st ACM SIGPLAN International Conference on Programming Language Design and Implementation, PLDI 2020, London, UK, June 15-20, 2020. pp. 703–717. ACM (2020). https://doi.org/10.1145/3385412.3386018, https://doi.org/10.1145/3385412.3386018
23. Koenig, J.R., Padon, O., Shoham, S., Aiken, A.: Inferring invariants with quantifier alternations: Taming the search space explosion. In: Fisman, D., Rosu, G. (eds.) Tools and Algorithms for the Construction and Analysis of Systems - 28th International Conference, TACAS 2022, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2022, Munich, Germany, April 2-7, 2022, Proceedings, Part I. Lecture Notes in Computer Science, vol. 13243, pp. 338–356. Springer (2022). https://doi.org/10.1007/978-3-030-99524-9_18, https://doi.org/10.1007/978-3-030-99524-9_18

24. Konnov, I., Kukovec, J., Tran, T.: TLA+ model checking made symbolic. Proc. ACM Program. Lang. **3**(OOPSLA), 123:1–123:30 (2019). https://doi.org/10.1145/3360549, https://doi.org/10.1145/3360549
25. Lamport, L.: The part-time parliament. ACM Transactions on Computer Systems **16**(2), 133–169 (1998)
26. Lamport, L.: Specifying Systems: The TLA$^+$ Language and Tools for Hardware and Software Engineers. Addison-Wesley Professional (Jul 2002)
27. Leino, K.R.M.: Dafny: An automatic program verifier for functional correctness. In: LPAR. LNCS, vol. 6355, pp. 348–370. Springer (2010)
28. Leino, K.R.M., Pit-Claudel, C.: Trigger selection strategies to stabilize program verifiers. In: Chaudhuri, S., Farzan, A. (eds.) Computer Aided Verification - 28th International Conference, CAV 2016, Toronto, ON, Canada, July 17-23, 2016, Proceedings, Part I. Lecture Notes in Computer Science, vol. 9779, pp. 361–381. Springer (2016). https://doi.org/10.1007/978-3-319-41528-4_20, https://doi.org/10.1007/978-3-319-41528-4_20
29. Macedo, N., Brunel, J., Chemouil, D., Cunha, A., Kuperberg, D.: Lightweight specification and analysis of dynamic systems with rich configurations. In: Zimmermann, T., Cleland-Huang, J., Su, Z. (eds.) Proceedings of the 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering, FSE 2016, Seattle, WA, USA, November 13-18, 2016. pp. 373–383. ACM (2016). https://doi.org/10.1145/2950290.2950318, https://doi.org/10.1145/2950290.2950318
30. McMillan, K.L., Padon, O.: Ivy: A multi-modal verification tool for distributed algorithms. In: Lahiri, S.K., Wang, C. (eds.) Computer Aided Verification - 32nd International Conference, CAV 2020, Los Angeles, CA, USA, July 21-24, 2020, Proceedings, Part II. Lecture Notes in Computer Science, vol. 12225, pp. 190–202. Springer (2020). https://doi.org/10.1007/978-3-030-53291-8_12, https://doi.org/10.1007/978-3-030-53291-8_12
31. McMillan, K.L., Padon, O.: Ivy: A multi-modal verification tool for distributed algorithms. In: Lahiri, S.K., Wang, C. (eds.) Computer Aided Verification - 32nd International Conference, CAV 2020, Los Angeles, CA, USA, July 21-24, 2020, Proceedings, Part II. Lecture Notes in Computer Science, vol. 12225, pp. 190–202. Springer (2020). https://doi.org/10.1007/978-3-030-53291-8_12, https://doi.org/10.1007/978-3-030-53291-8_12
32. Padon, O.: Deductive Verification of Distributed Protocols in First-Order Logic. Ph.D. thesis, Tel Aviv University (Dec 2018)
33. Padon, O., Losa, G., Sagiv, M., Shoham, S.: Paxos made EPR: decidable reasoning about distributed protocols. PACMPL **1**(OOPSLA), 108:1–108:31 (2017)
34. Padon, O., McMillan, K.L., Panda, A., Sagiv, M., Shoham, S.: Ivy: safety verification by interactive generalization. In: Proceedings of the 2016 ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI). pp. 614–630. Santa Barbara, CA (Jun 2016)
35. Padon, O., Wilcox, J.R., Koenig, J.R., McMillan, K.L., Aiken, A.: Induction duality: Primal-dual search for invariants. Proc. ACM Program. Lang. **6**(POPL), 1–29 (2022). https://doi.org/10.1145/3498712, https://doi.org/10.1145/3498712
36. Piskac, R., de Moura, L.M., Bjørner, N.S.: Deciding effectively propositional logic using DPLL and substitution sets. J. Autom. Reason. **44**(4), 401–424 (2010)
37. Pîrlea, G.: Translation from ivy to mypyvy (2024), https://github.com/kenmcmil/ivy/pull/76
38. Ramsey, F.P.: On a problem of formal logic. Proceedings of the London Mathematical Society **s2-30**(1), 264–286 (1930). https://doi.org/https:

//doi.org/10.1112/plms/s2-30.1.264, https://londmathsoc.onlinelibrary.wiley.com/doi/abs/10.1112/plms/s2-30.1.264

39. Taube, M., Losa, G., McMillan, K.L., Padon, O., Sagiv, M., Shoham, S., Wilcox, J.R., Woos, D.: Modularity for decidability of deductive verification with applications to distributed systems. In: Proceedings of the 2018 ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI). Philadelphia, PA (Jun 2018)
40. Yao, J., Tao, R., Gu, R., Nieh, J.: Duoai: Fast, automated inference of inductive invariants for verifying distributed protocols. In: Aguilera, M.K., Weatherspoon, H. (eds.) 16th USENIX Symposium on Operating Systems Design and Implementation, OSDI 2022, Carlsbad, CA, USA, July 11-13, 2022. pp. 485–501. USENIX Association (2022), https://www.usenix.org/conference/osdi22/presentation/yao
41. Yao, J., Tao, R., Gu, R., Nieh, J.: Mostly automated verification of liveness properties for distributed protocols with ranking functions. Proc. ACM Program. Lang. **8**(POPL) (jan 2024). https://doi.org/10.1145/3632877, https://doi.org/10.1145/3632877
42. Yao, J., Tao, R., Gu, R., Nieh, J., Jana, S., Ryan, G.: Distai: Data-driven automated invariant learning for distributed protocols. In: Brown, A.D., Lorch, J.R. (eds.) 15th USENIX Symposium on Operating Systems Design and Implementation, OSDI 2021, July 14-16, 2021. pp. 405–421. USENIX Association (2021), https://www.usenix.org/conference/osdi21/presentation/yao
43. Yu, Y., Manolios, P., Lamport, L.: Model checking TLA$^+$ specifications. In: Pierre, L., Kropf, T. (eds.) Correct Hardware Design and Verification Methods, 10th IFIP WG 10.5 Advanced Research Working Conference, CHARME '99, Bad Herrenalb, Germany, September 27-29, 1999, Proceedings. Lecture Notes in Computer Science, vol. 1703, pp. 54–66. Springer (1999)

A Examples of `mypyvy`'s output

When invoked on the toy consensus example, relaxed BMC of bound (depth) 5 returns the relaxed trace depicted in Fig. 4, indicating that universally quantified invariants are insufficient to prove safety in this case. The trace starts in a state whose universe has two nodes `node0`, `node1`, two values `value0`, `value1` and two quorums: `quorum0` = {`node1`}, `quorum1` = {`node0`, `node1`} (note that `quorum0` is not a majority set, but the quorum intersection axiom is satisfied). The relaxed trace consists of the transitions

1. vote(`node1`, `value0`),
2. decide(`value0`) based on quorum `quorum0`,
3. a relaxation step that omits node `node1` and quorum `quorum0` (as indicated by the fact that `active_node` does not hold for them after the `decrease_domain` step),
4. vote(`node0`, `value1`),
5. decide(`value1`) based on quorum `quorum1`, which after the relaxation step is just {`node0`} (the quorum intersection axioms are still satisfied since `quorum0` is gone).

At this point, both `value0` and `value1` have been decided, violating the agreement property.

```
> mypyvy bmc --depth=5 --relax consensus.pyv
bmc checking the following property up to depth 5
  forall X:value, Y:value. d(X) & d(Y) -> X = Y

found violation:
  universes:
    sort node
      node0
      node1
    sort quorum
      quorum0
      quorum1
    sort value
      value0
      value1

  immutable:
    member(node0,quorum1)
    member(node1,quorum0)
    member(node1,quorum1)

  state 0:
    active_node(node0)
    active_node(node1)
    active_quorum(quorum0)
    active_quorum(quorum1)
    active_value(value0)
    active_value(value1)

  transition vote

  state 1:
    active_node(node0)
    active_node(node1)
    active_quorum(quorum0)
    active_quorum(quorum1)
    active_value(value0)
    active_value(value1)
    b(node1)
    v(node1,value0)

  transition decide

  state 2:
    active_node(node0)
    active_node(node1)
    active_quorum(quorum0)
    active_quorum(quorum1)
    active_value(value0)
    active_value(value1)
    b(node1)
    d(value0)
    v(node1,value0)

  transition decrease_domain

  state 3:
    active_node(node0)
    active_quorum(quorum1)
    active_value(value0)
    active_value(value1)
    b(node1)
    d(value0)
    v(node1,value0)

  transition vote

  state 4:
    active_node(node0)
    active_quorum(quorum1)
    active_value(value0)
    active_value(value1)
    b(node0)
    d(value0)
    v(node0,value1)

  transition decide

  state 5:
    active_node(node0)
    active_quorum(quorum1)
    active_value(value0)
    active_value(value1)
    b(node0)
    d(value0)
    d(value1)
    v(node0,value1)
```

Fig. 4. A relaxed trace proving that the safety property of the toy consensus protocol cannot be proven with a universally quantified inductive invariant.

```
> mypyvy verify consensus.pyv

checking init:
  implies invariant agreement... ok.
checking transition vote:
  preserves invariant agreement... ok.
checking transition decide:
  preserves invariant agreement... no!

counterexample:
  universes:
    sort node
      node0
    sort quorum
      {node0}
    sort value
      value0
      value1

  immutable:
    member(node0,{node0})

  state 0:
    d(value1)
    v(node0,value0)

  state 1:
    d(value0)
    d(value1)
    v(node0,value0)

error consensus.pyv: invariant agreement is not preserved by transition decide
```

Fig. 5. A counterexample to induction (CTI) for the toy consensus protocol's safety property without additional invariants, printed using the custom set printer (`sort quorum @printed_by(set_printer, member)`).