Scavenger 0.1: A Theorem Prover Based on Conflict Resolution

Daniyar Itegulov, John Slaney, and Bruno Woltzenlogel Paleo*

[1] ITMO University, Russia
ditegulov@gmail.com
[2] Australian National University
john.slaney@anu.edu.au
bruno.wp@gmail.com

Abstract. This paper introduces Scavenger, the first theorem prover for pure first-order logic without equality based on the new conflict resolution calculus. Conflict resolution has a restricted resolution inference rule that resembles (a first-order generalization of) unit propagation as well as a rule for assuming decision literals and a rule for deriving new clauses by (a first-order generalization of) conflict-driven clause learning.

1 Introduction

The outstanding efficiency of current propositional SAT-solvers naturally raises the question of whether it would be possible to employ similar ideas for automating first-order logical reasoning. The recent *Conflict Resolution* calculus (**CR**) [21] can be regarded as a crucial initial step to answer this question. From a proof-theoretical perspective, it generalizes to first-order logic the two main mechanisms on which modern SAT-solvers are based: unit propagation and conflict-driven clause learning. The calculus is proven sound (by simulation by a clausal natural deduction calculus) and refutationally complete (by simulation of the usual resolution calculus), and it is shown that subderivations in **CR** are isomorphic to the implication graphs maintained by SAT-solvers.

This paper goes one step further by defining proof search algorithms for **CR**. Familiarity with the propositional CDCL procedure [14] is assumed, even though it is briefly sketched in Section 2. The main challenge in lifting this procedure to first-order logic is that, unlike in propositional logic, first-order unit propagation does not always terminate and true clauses do not necessarily have uniformly true literals (cf. Section 4). Our solutions to these challenges are discussed in Section 5 and Section 6, and experimental results are presented in Section 7.

Related Work: Other attempts to lift DPLL [8, 15] or CDCL [14] to first-order logic include *Model Evolution* [2, 4, 3, 20], *Geometric Resolution* [19], *Non-Redundant Clause Learning* [1] and the *Semantically-Guided Goal Sensitive procedure* [13]. A brief summary of these approaches and a comparison with **CR** can be found

* Author order is alphabetical by surname.

in [21]. In the terminology of [13], all these approaches can be classified as fine-grained with respect to the interleaving of proof search and model construction. In contrast, *Semantic Resolution* [22, 10] is an example of coarse-grained approach. Furthermore, many architectures [7, 11, 12, 25, 6] for first-order and higher-order theorem proving use a SAT-solver as a black box for propositional reasoning, without attempting to lift it.

2 Propositional CDCL

During search in the propositional case, a SAT-solver keeps a model (a.k.a. trail) consisting of a (conjunctive) list of decision literals and propagated literals. Literals of unit clauses are automatically added to the trail, and whenever a clause has only one literal that is not falsified by the current model, this literal is added to the model (thereby satisfying that clause). This process is known as *unit-propagation*. If unit propagation reaches a conflict (i.e. a situation where the dual of a literal already contained in the model would have to be added to it), the SAT-solver backtracks, removing from the model decision literals responsible for the conflict (as well as propagated literals entailed by the removed decision literals) and deriving, or learning, a conflict-driven clause consisting[1] of duals of the decision literals responsible for the conflict (or the empty clause, if there were no decision literals). If unit propagation terminates without reaching a conflict and all clauses are satisfied by the model, then the input clause set is satisfiable. If some clauses are still not satisfied, the SAT-solver chooses and assigns another decision literal, adding it to the trail, and satisfying the clauses that contain it.

3 Conflict Resolution

The inference rules of the conflict resolution calculus **CR** are shown in Fig. 1. The unit propagating resolution rule is a chain of restricted resolutions with unit clauses as left premises and a unit clause as final conclusion. Decision literals are denoted by square brackets, and the conflict-driven clause learning rule allows to infer a new clause consisting of negations of instances of decision literals used to reach a conflict (a.k.a. the empty clause $\bot$). A clause learning inference is said to discharge the decision literals that it uses. As in the resolution calculus, **CR** derivations are directed acyclic graphs that are not necessarily tree-like. A **CR** refutation is a **CR** derivation of $\bot$ with no undischarged decision literals.

From a natural deduction point of view, a unit propagating resolution rule can be regarded as a chain of implication eliminations taking unification into account, whereas decision literals and conflict driven clause learning are reminiscent of, respectively, assumptions and chains of negation introductions, also generalized to first-order through unification. Therefore, **CR** can be considered a first-order hybrid of resolution and natural deduction.

[1] In practice, optimizations (e.g. 1UIP) are used, and more sophisticated clauses, which are not just disjunctions of duals of the decision literals involved in the conflict, can be derived. But these optimizations are inessential to the focus of this paper.

Unit-Propagating Resolution:

$$\frac{\ell_1 \quad \ldots \quad \ell_n \quad \overline{\ell'_1} \vee \ldots \vee \overline{\ell'_n} \vee \ell}{\ell\ \sigma}\ \mathbf{u}(\sigma)$$

where σ is a unifier of ℓ_k and ℓ'_k, for all $k \in \{1, \ldots, n\}$.

Conflict:

$$\frac{\ell \quad \overline{\ell'}}{\bot}\ \mathbf{c}(\sigma)$$

where σ is a unifier of ℓ and ℓ'.

Conflict-Driven Clause Learning:

$$\frac{\begin{array}{cc}[\ell_1]^i_1 & [\ell_n]^i_n \\ \vdots\ (\sigma^1_1, \ldots, \sigma^1_{m_1}) & \vdots\ (\sigma^n_1, \ldots, \sigma^n_{m_n}) \\ & \vdots \\ & \bot\end{array}}{(\overline{\ell_1}\sigma^1_1 \vee \ldots \vee \overline{\ell_1}\sigma^1_{m_1}) \vee \ldots \vee (\overline{\ell_n}\sigma^n_1 \vee \ldots \vee \overline{\ell_n}\sigma^n_{m_n})}\ \mathbf{cl}^i$$

where σ^k_j (for $1 \leq k \leq n$ and $1 \leq j \leq m_k$) is the composition of all substitutions used on the j-th path[a] from ℓ_k to $\bot$.

[a] Since a proof DAG is not necessarily tree-like, there may be more than one path connecting ℓ_k to $\bot$ in the DAG-like proof.

Fig. 1: The Conflict Resolution Calculus **CR**

4 Lifting Challenges

First-order logic presents many new challenges for methods based on propagation and decisions, of which the following can be singled out:

(1) *non-termination of unit-propagation:* In first-order logic, unit propagation may never terminate. For example, the clause set $\{p(a), \neg p(X) \vee p(f(X)), q \vee r, \neg q \vee r, q \vee \neg r, \neg q \vee \neg r\}$ is clearly unsatisfiable, because there is no assignment of p and q to *true* or *false* that would satisfy all the last four clauses. However, unit propagation would derive the following infinite sequence of units, by successively resolving $\neg p(X) \vee p(f(X))$ with previously derived units, starting with $p(a)$: $\{p(f(a)), p(f(f(a))), \ldots, p(f(\ldots(f(a))\ldots)), \ldots\}$. Consequently, a proof search strategy that would wait for unit propagation to terminate before making decisions would never be able to conclude that the given clause set is unsatisfiable.

(2) *absence of uniformly true literals in satisfied clauses:* While in the propositional case, a clause that is true in a model always has at least one literal

that is true in that model, this is not so in first-order logic, because shared variables create dependencies between literals. For instance, the clause set $\{p(X) \vee q(X), \neg p(a), p(b), q(a), \neg q(b)\}$ is satisfiable, but there is no model where $p(X)$ is uniformly true (i.e. true for all instances of X) or $q(X)$ is uniformly true.

(3) *propagation without satisfaction:* In the propositional case, when only one literal of a clause is not false in the model, this literal is propagated and added to the model, and the clause necessarily becomes true in the model and does not need to be considered in propagation anymore, at least until backtracking. In the first-order case, on the other hand, a clause such as $p(X) \vee q(X)$ would propagate the literal $q(a)$ in a model containing $\neg p(a)$, but $p(X) \vee q(X)$ does not become true in a model where $q(a)$ is true. It must remain available for further propagations. If, for instance, the literal $\neg p(b)$ is added to the model, the clause will be used again to propagate $q(b)$.

(4) *quasi-falsification without propagation:* A clause is *quasi-falsified* by a model iff all but one of its literals are false in the model. In first-order logic, in contrast to propositional logic, it is not even the case that a clause will necessarily propagate a literal when only one of its literals is not false in the model. For instance, the clause $p(X) \vee q(X) \vee r(X)$ is quasi-falsified in a model containing $\neg p(a)$ and $\neg q(b)$, but no instance of $r(X)$ can be propagated.

The first two challenges affect the search in a conceptual level, and possible solutions are discussed in Section 5. The last two challenges prevent a direct first-order generalization of the data structures (e.g. *watched literals*) that make unit propagation so efficient in the propositional case. Partial solutions are discussed in Section 6.

5 First-Order Model Construction and Proof Search

Despite the fundamental differences between propositional and first-order logic described in the previous section, the first-order algorithms presented aim to adhere as much as possible to the propositional procedure sketched in the Section 2. As in the propositional case, the model under construction is a (conjunctive) list of literals, but literals may now contain (universal) variables. If a literal $\ell[X]$ is in a model M, then any instance $\ell[t]$ is said to be true in M. Note that checking that a literal ℓ is true in a model M is more expensive in first-order logic than in propositional logic: whereas in the latter it suffices to check that ℓ is in M, in the former it is necessary to find a literal ℓ' in M and a substitution σ such that $\ell = \ell'\sigma$. A literal ℓ is said to be *strongly true* in a model M iff ℓ is in M.

There is a straightforward solution for the second challenge (i.e. the absence of uniformly true literals in satisfied clauses): a clause is satisfied by a model M iff all its relevant instances have a literal that is true in M, where an instance is said to be relevant if it substitutes the clause's variables by terms that occur in M. Thus, for instance, the clause $p(X) \vee q(X)$ is satisfied by the model $[\neg p(a), p(b), q(a), \neg q(b)]$, because both relevant instances $p(a) \vee q(a)$ and $p(b) \vee q(b)$ have literals that are

true in the model. However, this solution is clearly costly, because it requires the generation of many instances. Fortunately, in many (though not all) cases, a satisfied clause will have a literal that is true in M, in which case the clause is said to be *uniformly satisfied*. Checking uniform satisfaction is cheaper than checking satisfaction. However, a drawback of uniform satisfaction is that the model construction algorithm may repeatedly attempt to satisfy a clause that is not uniformly satisfied, by choosing one of its literals as a decision literal. For instance, the clause $p(X) \vee q(X)$ is not uniformly satisfied by the model $[\neg p(a), p(b), q(a), \neg q(b)]$. Without knowing that this clause is already satisfied by the model, the procedure would try to choose either $p(X)$ or $q(X)$ as a decision literal. But any of these choices is *useless decision*, because they would lead to a conflict with resulting conflict-driven clause equal to a previously derived clause or to a unit clause containing a literal that is part of the current model. A clause is said to be *weakly satisfied* by a model M if and only if all its literals are useless decisions.

Because of the first challenge (i.e. the non-termination of unit-propagation in the general first-order case), it is crucial to make decisions *during* unit propagation. In the given example in the previous section, for instance, deciding q at any moment would allow the propagation of r and $\neg r$ (respectively due to the 4th and 6th clauses), triggering a conflict. The learned clause would be $\neg q$ and it would again trigger a conflict by the propagation of r and $\neg r$ (this time due to the 3rd and 5th clauses). As this last conflict does not depend on any decision literal, the empty clause is derived and thus the clause set is refuted.

The question is how to interleave decisions and propagations. One straightforward approach is to keep track of the *propagation depth*[2] in the implication graph: any decision literal or literal propagated by a unit clause has propagation depth 0; any other literal has propagation depth $k + 1$, where k is the maximum propagation depth of its predecessors. Then propagation is performed exhaustively only up to a propagation depth threshold h. A decision literal is then chosen and the threshold is incremented. Such *eager decisions* guarantee that a decision will eventually be made, even if there is an infinite propagation path. However, eager decisions may also lead to spurious conflicts generating useless conflict-driven clauses. For instance, the clause set $\{1 : p(a), 2 : \neg p(X) \vee p(f(X)), 3 : \neg p(f(f(f(f(f(f(a))))))), 4 : \neg r(X) \vee q(X), 5 : \neg q(g(X)) \vee \neg p(X), 6 : z(X) \vee r(X)\}$ (where clauses have been numbered for easier reference) is unsatisfiable, because a conflict with no decisions can be obtained by propagating $p(a)$ (by 1), and then $p(f(a))$, $p(f(f(a)))$, $\ldots$, $p(f(f(f(f(f(f(a)))))))$, (by 2, repeatedly), which conflicts with $\neg p(f(f(f(f(f(f(a)))))))$ (by 3). But the former propagation has depth 6. If the propagation depth threshold is lower than 6, a decision literal is chosen before that conflict is reached. If $r(X)$ is chosen, for example, in an

[2] Because of the isomorphism between implication graphs and subderivations in Conflict Resolution [21], the propagation depth is equal to the corresponding subderivation's *height*, where initial axiom clauses and learned clauses have height 0 and the height of the conclusion of a unit-propagating resolution inference is $k + 1$ where k is the maximum height of its unit premises.

attempt to satisfy the sixth clause, there are propagations (using $r(X)$ and clauses 1, 4, 5 and 6) with depth lower than the threshold and reaching a conflict that generates the clause $\neg r(g(a))$, which is useless to show unsatisfiability of the whole clause set. This is not a serious issue, because useless clauses are often generated in conflicts with non-eager decisions as well. Nevertheless, this example suggests that the starting threshold and the strategy for increasing the threshold have to be chosen wisely, since the performance may be sensitive to this choice.

Interestingly, the problem of non-terminating propagation does not manifest in fragments of first-order logic where infinite unit propagation paths are impossible. A well-known and large fragment is the *effectively propositional* (a.k.a. *Bernays-Schönfinkel*) class, consisting of sentences with prenex forms that have an $\exists^*\forall^*$ quantifier prefix and no function symbols. For this fragment, a simpler proof search strategy that only makes decisions when unit propagation terminates, as in the propositional case, suffices. Infinite unit propagation paths do not occur in the effectively propositional fragment because there are no function symbols and hence the term depth[3] does not increase arbitrarily. Whenever the term depth is bounded, infinite unit propagation paths cannot occur, because there are only finitely many literals with bounded term depth (given the finite set of constant, function and predicate symbols with finite arity occurring in the clause set).

The insight that term depth is important naturally suggests a different approach for the general first-order case: instead of limiting the propagation depth, limit the *term depth* instead, allowing arbitrarily long propagations as long as the term depth of the propagated literals are smaller than the current term depth threshold. A literal is propagated only if its term depth is smaller than the threshold. New decisions are chosen when the term-depth-bounded propagation terminates and there are still clauses that are not uniformly satisfied. As before, eager decisions may lead to spurious conflicts, but bounding propagation by term depth seems intuitively more sensible than bounding it by propagation depth.

6 Implementation Details

Scavenger is implemented in Scala and its source code and usage instructions are available in `https://gitlab.com/aossie/Scavenger`. Its packrat combinator parsers are able to parse TPTP CNF files without *let* expressions [24]. Although Scavenger is a first-order prover, every logical expression is converted to a simply typed lambda expression, implemented by the abstract class `E` with concrete subclasses `Sym`, `App` and `Abs` for, respectively, *symbols*, *applications* and *abstractions*. A trait `Var` is used to distinguish *variables* from other symbols. Scala's *case classes* are used to make `E` behave like an algebraic datatype with (pattern-matchable) constructors. Simply typed lambda expressions are chosen despite Scavenger's current focus on untyped first-order logic, because we intend to generalize Scavenger to multi-sorted first-order logic and higher-order logic and support TPTP TFF and THF in the future. Every clause is internally represented

[3] The depth of constants and variables is zero and the depth of a complex term is $k+1$ when k is the maximum depth of its proper subterms.

as an immutable two-sided sequent consisting of a set of positive literals in the succedent and a set of negative literals in the antecedent.

When a problem is unsatisfiable, Scavenger can output a **CR** refutation, which is internally represented as a collection of `ProofNode` objects, which can be instances of the following immutable classes: `UnitPropagatingResolution`, `Conflict`, `ConflictDrivenClauseLearning`, `Axiom`, `Decision`. The first three classes correspond directly to the rules shown in Fig. 1. `Axiom` is used for leaf nodes containing input clauses, and `Decision` represents a fictive rule holding decision literals. Each class is responsible for checking, typically through `require` statements, the soundness conditions of its corresponding inference rule. The `Axiom`, `Decision` and `ConflictDrivenClauseLearning` classes are less than 5 lines of code each. `Conflict` and `UnitPropagatingResolution` are respectively 15 and 35 lines of code. The code for analyzing conflicts, traversing the subderivations (conflict graphs) and finding decisions that contributed to the conflict, is implemented in a superclass, and is 17 lines long.

The following three variants of Scavenger were implemented:

- EP-Scavenger: This variant aims at the effectively propositional fragment. Propagation is not bounded, and decisions are made only when propagation terminates.
- PD-Scavenger: Propagation is bounded by a propagation depth threshold starting at 0. Input clauses are assigned depth 0. Derived clauses and propagated literals obtained while the depth threshold is k are assigned depth $k+1$. The threshold is incremented whenever every input clause that is neither uniformly satisfied nor weakly satisfied is used to derive a new clause or to propagate a new literal. If this is not the case, a decision literal is chosen (and assigned depth $k+1$) to uniformly satisfy one of the clauses that is neither uniformly satisfied nor weakly satisfied.
- TD-Scavenger: Propagation is bounded by a term depth threshold starting at 0 and incrementing with 50% probability whenever propagation terminates (and choosing a decision literal when the threshold is not incremented). Only uniform satisfaction of clauses is checked.

The third and fourth challenges discussed in Section 4 are critical for performance, because they prevent a direct first-order generalization of data structures such as *watched literals*, which enables efficient detection of clauses that are ready to propagate literals. Without knowing exactly which clauses are ready to propagate, Scavenger (in its three variants) loops through all clauses with the goal of using them for propagation. However, actually trying to use a given clause for propagation is costly. In order to avoid this cost, Scavenger performs two quicker tests. Firstly, it checks whether the clause is uniformly satisfied (by checking whether one of its literals belongs to the model). If it is, then the clause is dismissed. This is an imperfect test, however. Occasionally, some satisfied clauses will not be dismissed, because (in first-order logic) not all satisfied clauses are uniformly satisfied. Secondly, for every literal ℓ of every clause, Scavenger keeps a set of decision literals and propagated literals that are unifiable with ℓ. A clause c is quasi-falsified when at most one literal of c has an empty set associated with it.

This is a rough analogue of watched literals for detecting quasi-falsified clauses. Again, this is an imperfect test, because (in first-order logic) not all quasi-falsified clauses are ready to propagate. Despite the imperfections of these tests, they do reduce the number of clauses that need to be considered for propagation, and they are quick and simple to implement.

Overall, the three variants of Scavenger listed above have been implemented very concisely. Their main classes are only 168, 342 and 176 lines long, respectively, and no attempt has been made to increase efficiency at the expense of code readability.

Scavenger still has no sophisticated backtracking and restarting mechanism, as propositional SAT-solvers do. When Scavenger reaches a conflict, it restarts almost completely: all derived conflict-driven clauses are kept, but the model under construction is reset to the empty model.

7 Experiments

Experiments were conducted[4] in the StarExec cluster [23] to evaluate Scavenger's performance on TPTP v6.4.0 benchmarks in CNF form and without equality. For comparison, all 29 provers available in StarExec's TPTP community and capable of reasoning on the selected benchmarks were evaluated as well. For each job pair, the timeouts were 300 CPU seconds and 600 Wallclock seconds.

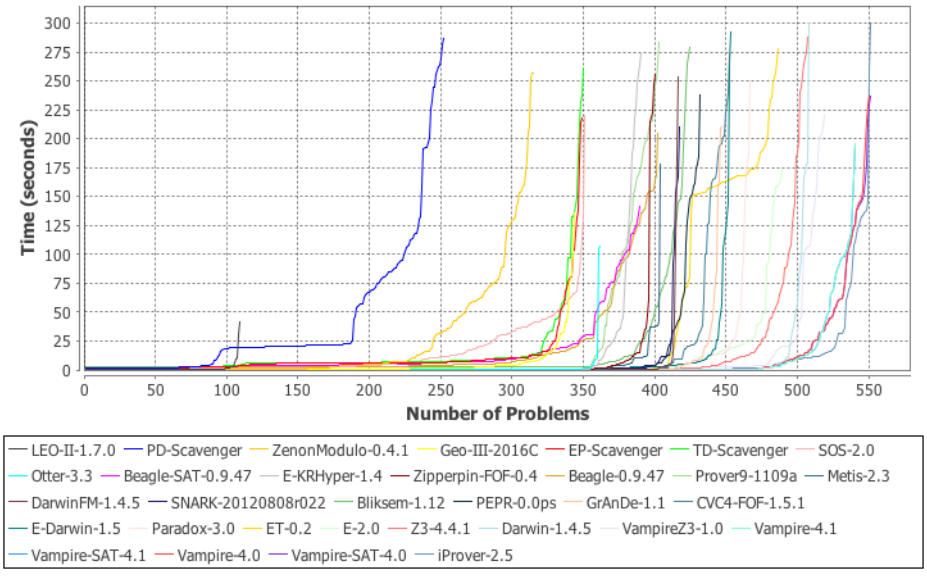

Fig. 2: Performance on Unsat. Eff. Propositional CNF Problems without Equality

[4] Raw experimental data are available at `https://doi.org/10.5281/zenodo.293187`.

Fig 2 shows how many of the 572 unsatisfiable effectively propositional problems each prover can solve within a given amount of time. As expected, TD-Scavenger outperforms PD-Scavenger, supporting the intuition that term depth is a more natural criterion for bounding unit propagation, and EP-Scavenger tends to be slightly faster than TD-Scavenger, although TD-Scavenger surprisingly solved one problem more than EP-Scavenger. For a first implementation, the best variants of Scavenger show an acceptable performance. Capable of solving 350 problems within the 300s time limit, TD-Scavenger outperformed LEO-II, ZenonModulo and Geo-III, and solved only 1 problem less than SOS-2.0 and 12 less than Otter-3.3. Although Otter-3.3 has long ceased to be a state-of-the-art prover and has been replaced by Prover9, the fact that Scavenger solves almost as many problems as Otter-3.3 is encouraging, because Otter-3.3 is a mature prover with 15 years of development, implementing (in the C language) several refinements of proof search for resolution and paramodulation (e.g. orderings, set of support, splitting, demodulation, subsumption) [16, 17], whereas Scavenger is a yet unrefined and concise implementation (in Scala) of a comparatively straightforward search strategy for proofs in the Conflict Resolution calculus, completed in slightly more than 3 months.

Figure 3 shows the performance on all 1606 unsatisfiable (not necessarily effectively propositional) problems. All variants of Scavenger outperformed PEPR, GrAnDe, DarwinFM, Paradox, ZenonModulo and LEO-II; and EP-Scavenger additionally outperformed Geo-III. Solving 891 problems, EP-Scavenger was significantly better than PD-Scavenger (782) and TD-Scavenger (695). This suggests that non-termination of unit-propagation is an uncommon issue in practice: EP-Scavenger is still able to solve many problems, even though it does not care to bound propagation, whereas the other two variants solve fewer problems because of the overhead of bounding propagation even when it is not necessary. Nevertheless, there were 28 problems solved only by PD-Scavenger and 26 problems solved only by TD-Scavenger (among Scavenger's variants).

8 Conclusions and Future Work

Scavenger is the first theorem prover based on the new Conflict Resolution calculus. The experiments show that its performance is promising, albeit not yet competitive.

A comparison of the performance of the three variants of Scavenger shows that it is non-trivial to interleave decisions within possibly non-terminating unit-propagations, and further research is needed to determine (possibly in a problem dependent way) optimal initial depth thresholds and threshold incrementation strategies. Alternatively, entirely different criteria could be explored for deciding to make an eager decision before propagation is over. For instance, decisions could be made if a fixed or dynamically adjusted amount of time elapses.

The performance bottleneck that needs to be most urgently addressed in future work is backtracking and restarting. Currently, all variants of Scavenger restart after every conflict, keeping derived conflict-driven clauses but throwing away the

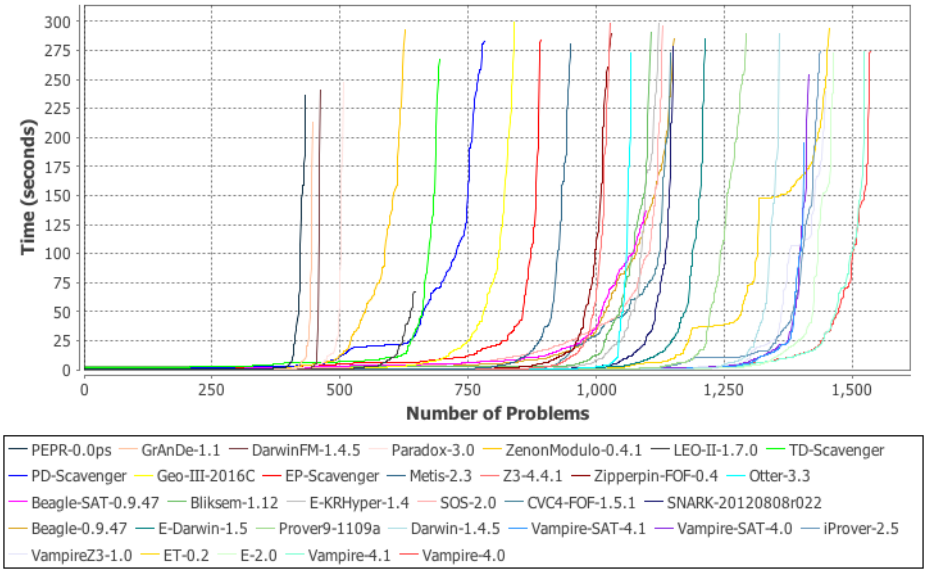

Fig. 3: Performance on Unsatisfiable CNF Problems without Equality

model construct so far. They must reconstruct models from scratch after every conflict. This requires a lot of repeated re-computation, and therefore a significant performance boost could be expected through a more sensible backtracking strategy. There might also be room to improve Scavenger's rough first-order analogue for the *watched literals* data structure, even though the first-order challenges make it unlikely that something as good as the propositional watched literals data structure could ever be developed. Furthermore Scavenger currently uses no term indexing [18] and its unification algorithm is implemented naively. Scavenger inherits from the proof compression system `Skeptik` [5] many data structures that had been implemented aiming at convenient proof manipulation instead of efficient theorem proving.

Scavenger's already acceptable performance despite the implementation improvement possibilities just discussed above indicates that automated theorem proving based on the Conflict Resolution calculus is feasible. However, much work remains to be done to determine whether this approach will eventually become competitive with today's fastest provers.

Acknowledgments: We thank Ezequiel Postan for his implementation of TPTP parsers for `Skeptik` [5], which we have reused in Scavenger. We thank Albert A. V. Giegerich, Aaron Stump and Geoff Sutcliffe for all their help in setting up our experiments in StarExec. This research was partially funded by the Australian Government through the Australian Research Council and by the Google Summer of Code 2016 program.

References

1. Alagi, G., Weidenbach, C.: Non-redundant clause learning. In: FroCoS. pp. 69–84 (2015)
2. Baumgartner, P.: A first order Davis-Putnam-Longeman-Loveland procedure. In: Proceedings of the 17th International Conference on Automated Deduction (CADE). pp. 200–219 (2000)
3. Baumgartner, P.: Model evolution based theorem proving. IEEE Inteligent Systems 29(1), 4–10 (2014)
4. Baumgartner, P., Tinelli, C.: The model evolution calculus. In: CADE. pp. 350–364 (2003)
5. Boudou, J., Fellner, A., Paleo, B.W.: Skeptik: A proof compression system. In: Demri, S., Kapur, D., Weidenbach, C. (eds.) Automated Reasoning - 7th International Joint Conference, IJCAR 2014, Held as Part of the Vienna Summer of Logic, VSL 2014, Vienna, Austria, July 19-22, 2014. Proceedings. Lecture Notes in Computer Science, vol. 8562, pp. 374–380. Springer (2014), http://dx.doi.org/10.1007/978-3-319-08587-6_29
6. Brown, C.E.: Satallax: An automatic higher-order prover. In: Gramlich, B., Miller, D., Sattler, U. (eds.) IJCAR. Lecture Notes in Computer Science, vol. 7364, pp. 111–117. Springer (2012)
7. Claessen, K.: The anatomy of Equinox – an extensible automated reasoning tool for first-order logic and beyond (talk abstract). In: Proceedings of the 23rd International Conference on Automated Deduction (CADE-23). pp. 1–3 (2011)
8. Davis, M., Putnam, H.: A computing procedure for quantification theory. Journal of the ACM 7, 201–215 (1960)
9. Goré, R., Leitsch, A., Nipkow, T. (eds.): Automated Reasoning, First International Joint Conference, IJCAR 2001, Siena, Italy, June 18-23, 2001, Proceedings, Lecture Notes in Computer Science, vol. 2083. Springer (2001)
10. Hodgson, K., Slaney, J.K.: System description: SCOTT-5. In: Goré et al. [9], pp. 443–447, http://dx.doi.org/10.1007/3-540-45744-5_36
11. Korovin, K.: iProver - an instantiation-based theorem prover for first-order logic (system description). In: Armando, A., Baumgartner, P., Dowek, G. (eds.) IJCAR. Lecture Notes in Computer Science, vol. 5195, pp. 292–298. Springer (2008)
12. Korovin, K.: Inst-Gen - a modular approach to instantiation-based automated reasoning. In: Programming Logics. pp. 239–270 (2013)
13. Maria Paola Bonacina, U.F., Sofronie-Stokkermans, V.: On first-order model-based reasoning. In: Logic, Rewriting and Concurrency. pp. 181–204 (2015)
14. João Marques-Silva, I.L., Malik, S.: Conflict-driven clause learning SAT solvers. In: Handbook of Satisfiability, pp. 127 – 149 (2008)
15. Martin Davis, G.L., Loveland, D.: A machine program for theorem proving. Communications of the ACM 57, 394–397 (1962)
16. McCune, W.: OTTER 2.0. In: Stickel, M.E. (ed.) 10th International Conference on Automated Deduction, Kaiserslautern, FRG, July 24-27, 1990, Proceedings. Lecture Notes in Computer Science, vol. 449, pp. 663–664. Springer (1990), http://dx.doi.org/10.1007/3-540-52885-7_131
17. McCune, W.: OTTER 3.3 reference manual. CoRR cs.SC/0310056 (2003), http://arxiv.org/abs/cs.SC/0310056
18. Nieuwenhuis, R., Hillenbrand, T., Riazanov, A., Voronkov, A.: On the evaluation of indexing techniques for theorem proving. In: Goré et al. [9], pp. 257–271, http://dx.doi.org/10.1007/3-540-45744-5_19

19. de Nivelle, H., Meng, J.: Geometric resolution: A proof procedure based on finite model search. In: 3rd International Joint Conference on Automated Reasoning (IJCAR). pp. 303–317 (2006)
20. Peter Baumgartner, A.F., Tinelli, C.: Lemma learning in the model evolution calculus. In: LPAR. pp. 572–586 (2006)
21. Slaney, J., Woltzenlogel Paleo, B.: Conflict Resolution: a first-order resolution calculus with decision literals and conflict-driven clause learning. Submitted. (Preprint: https://arxiv.org/pdf/1602.04568.pdf) (2016)
22. Slaney, J.K.: SCOTT: A model-guided theorem prover. In: Bajcsy, R. (ed.) Proceedings of the 13th International Joint Conference on Artificial Intelligence. Chambéry, France, August 28 - September 3, 1993. pp. 109–115. Morgan Kaufmann (1993), http://ijcai.org/Proceedings/93-1/Papers/016.pdf
23. Stump, A., Sutcliffe, G., Tinelli, C.: StarExec: A cross-community infrastructure for logic solving. In: Demri, S., Kapur, D., Weidenbach, C. (eds.) Automated Reasoning: 7th International Joint Conference, IJCAR 2014, Held as Part of the Vienna Summer of Logic, VSL 2014, Vienna, Austria, July 19-22, 2014. Proceedings. pp. 367–373. Springer International Publishing, Cham (2014), http://dx.doi.org/10.1007/978-3-319-08587-6_28
24. Sutcliffe, G.: The TPTP problem library and associated infrastructure: The FOF and CNF parts, v3.5.0. Journal of Automated Reasoning 43(4), 337–362 (2009)
25. Voronkov, A.: AVATAR: The architecture for first-order theorem provers. In: CAV. pp. 696–710 (2014)