Abstract

Datalog has become a popular language for writing static analyses. Because Datalog is very limited, some implementations of Datalog for static analysis have extended it with new language features. However, even with these features it is hard or impossible to express a large class of analyses because they use logical formulae to represent program state. FormuLog fills this gap by extending Datalog to represent, manipulate, and reason about logical formulae. We have used FormuLog to implement declarative versions of symbolic execution and abstract model checking, analyses previously out of the scope of Datalog-based languages.

While this paper focuses on the design of FormuLog and one of the analyses we have implemented in it, it also touches on a prototype implementation of the language and identifies performance optimizations that we believe will be necessary to scale FormuLog to real-world static analysis problems.

KEYWORDS: Datalog, static analysis, symbolic execution, model checking

1 Introduction

The logic programming language Datalog has become a popular domain-specific language for writing static analyses (Reps 1995, Whaley et al. 2005, Bravenboer and Smaragdakis 2009, Scholz et al. 2016, Madsen et al. 2016). As a declarative language that embodies Kolwaski’s principle of separating a program’s “logic” (what it computes) from its “control” (how it computes it) (Kowalski 1979), Datalog is a good fit for programming a static analysis, where often the logic of the analysis is substantially simpler than the control necessary to efficiently compute it. In particular, many analyses logically consist of mutually recursive sub-analyses that in theory interact elegantly, but in practice can be tricky to coordinate. Datalog makes it easy to state the dependencies found in these types of analyses, enabling analysis designers to focus on perfecting the logic of their analyses without worrying about the low-level control details.

Because Datalog is a very restricted language and there are many analyses that cannot easily be expressed in pure Datalog, implementations of Datalog for static analysis have extended the language with additional features. For instance, Soufflé adds a form of n-ary constructors (Scholz et al. 2016), and Flix adds user-defined functions and support for reasoning over lattices (Madsen et al. 2016). However, there are still common analyses that cannot be naturally expressed even
with these features: namely, analyses that reason about program behavior through logical formulae. Such analyses include symbolic execution, abstract interpretation over the predicate domain, and various forms of model checking. These analyses symbolically represent reachable program states as logical formulae, and then reason about them in theories that support concepts such as machine integers and arrays. No Datalog variant currently provides language abstractions for computing over these types of formulae.

This paper presents a language called FormuLog that extends Datalog to support this class of analyses, and thus explores how the declarative benefits of Datalog can be extended to these analyses. FormuLog makes it easy to represent, manipulate, and reason about logical formulae:

- Logical formulae can be represented by ground (i.e., variable-less) terms formed from n-ary constructors, such as the term \( \text{or}(\text{true}, \text{false}) \), which encodes the formula \( \text{True} \lor \text{False} \) using the binary constructor \( \text{or} \) and the nullary constructors \( \text{true} \) and \( \text{false} \).
- Logical formulae can be manipulated by user-defined functions that consume and produce ground terms. For example, a function might rename the logical variables in a formula.
- Logical formulae can be reasoned about through built-in functions that invoke an external satisfiability modulo theories (SMT) solver. For instance, the unary function \( \text{is_sat} \) evaluates to true if its argument is satisfiable when interpreted as a logical formula. A type system ensures that only terms that can be interpreted as formulae are applied to these special functions.

These features are combined in such a way that FormuLog remains a declarative language, in that the meaning of a FormuLog program is independent of how it is evaluated. On the one hand, this means that analysis designers can reason about the correctness of their analyses without worrying about how the FormuLog runtime will actually execute them. On the other hand, this means that the FormuLog runtime is free to choose how to evaluate an analysis and can aggressively apply optimizations, such as rewriting the analysis and running it in parallel.

Languages like FormuLog can help address concrete challenges faced by modern static analyses. For example, Toman and Grossman (2017) explain how state-of-the-art static analyses struggle to analyze programs that make heavy use of frameworks, such as Java web applications. These type of applications are difficult to analyze because frameworks typically rely on dynamic language features like reflection. Analyses cannot ignore these features because they play such a large role in framework behavior, but most analyses that treat these features precisely do not scale. Because of this, Toman and Grossman suggest a hybrid approach: a scalable meet-over-all-paths analysis is used for the application code, and a variant of symbolic execution, being more precise but less scalable, is used for the framework code. As control flow goes back and forth between the application and the framework, these analyses interact in a mutually recursive way. This is the perfect setting for a Datalog-like language that makes it easy to encode interdependent analyses. But, of course, that language must be able to effectively support the relevant analyses, including analyses such as symbolic execution that are currently not supported by Datalog variants. FormuLog takes a step in this direction by making it easier to declaratively encode analyses that compute with logical formulae.

Section 2 provides background on Datalog and how logic programming has previously been used for static analysis. Section 3 presents the design of FormuLog and Section 4 demonstrates how it enables a declarative encoding of symbolic execution. Section 5 discusses a current prototype of the language, gives some initial performance results, and sketches potential optimizations.
FormuLog: Datalog for static analysis involving logical formulae

### Variables
- \( X, Y, Z \in \text{Var} \)

### Constructor symbol
- \( a, b \in \text{CSym} \)

### Predicate symbol
- \( p \in \text{PSym} \)

### Term
- \( s, t ::= X \mid a \)

### Atom
- \( A ::= p(t_1, \ldots, t_n) \)

### Clause
- \( C ::= A : - A_1, \ldots, A_n \).

**Fig. 1.** Datalog is a simple language that forms the basis of FormuLog.

## 2 Background and related work

Datalog is a simple logic programming language (Figure 1) ([Ceri et al. 1989](https://doi.org/10.2307/3677825) | [Green et al. 2013](https://doi.org/10.1109/ISVC.2013.230)). A Datalog program is defined by a set of clauses, where each clause consists of a single head atom and a set of body atoms. An atom is a predicate symbol applied to a list of terms, where each term is either a variable or an uninterpreted constant (i.e., a nullary constructor). Each predicate symbol is associated with either an extensional database (EDB) relation or an intensional database (IDB) relation. An EDB relation is tabulated explicitly through facts (clauses with empty bodies), whereas an IDB relation is computed through rules, clauses that have non-empty bodies.

Datalog has an elegant logical interpretation that leads to a declarative semantics, in the sense that the meaning of a Datalog program does not depend on how it is evaluated. A clause

\[ A : - A_1, \ldots, A_n. \]

can be interpreted as the logical formula

\[ \forall \overline{X}. (A_1 \land \cdots \land A_n \rightarrow A) \]

where \( \overline{X} \) is a vector of the variables that occur in the clause. The clauses of the program can then be viewed as the axioms of a first-order theory, and the semantics of the program is the least Herbrand model of that theory (Datalog puts syntactic restrictions on programs to guarantee that such a model exists and is finite).

### 2.1 Datalog-based static analysis frameworks

Subsequent to the early use of Datalog by [Reps (1995)](https://doi.org/10.1109/SFCS.1995.492117) for deriving demand-driven versions of interprocedural analyses, a variety of static analysis frameworks have been developed that leverage Datalog. A binary decision diagram-based implementation of Datalog, bddbddb ([Whaley et al. 2005](https://doi.org/10.1145/1066682.1066687)), has been used to compute context-sensitive pointer analysis ([Whaley and Lam 2004](https://doi.org/10.1145/1066682.1066687)) and taint analysis for Java applications ([Livshits and Lam 2005](https://doi.org/10.1145/1066682.1066687)), and has also been used as one of the analysis languages for Chord, a Java bytecode analysis framework. The points-to analysis framework Doop uses a heavily-optimized version of Datalog to outperform contemporary state-of-the-art pointer analyses built on binary decision diagrams ([Bravenboer and Smaragdakis 2009](https://doi.org/10.1109/ISVCS.2009.160)). Instead of evaluating an analysis written in Datalog directly, Soufflé treats the analysis as a synthesis specification: from the Datalog program, it synthesizes a C++ implementation of that analysis ([Scholz et al. 2016](https://doi.org/10.1145/2945622.2945648) | [Jordan et al. 2016](https://doi.org/10.1145/2945622.2945648)). Soufflé extends Datalog with a type system and a

---

1 Chord is available at [https://bitbucket.org/psl-lab/jchord/](https://bitbucket.org/psl-lab/jchord/)
form of \textit{n}-ary constructors, and has been used to synthesize pointer and security analyses for Java applications. Because the addition of \textit{n}-ary constructors makes Datalog Turing-complete (Green et al. 2013), Soufflé is technically powerful enough to encode the analyses that FormuLog is designed for; however, it does not provide the abstractions to make this practical.

Flix extends Datalog with monotone functions written in a pure functional language, algebraic data types, and user-defined lattices (Madsen et al. 2016). Although FormuLog also has pure functions and algebraic data types, the two languages target different shortcomings in the use of Datalog for static analysis. Flix addresses the fact that it is inefficient or impossible to state analyses in Datalog that compute over lattices that are not the powerset lattice. FormuLog addresses the fact that Datalog provides no way to manipulate and reason about logical state. That being said, FormuLog would benefit from some type of aggregation, and the lattice approach of Flix might be a good fit. A recent paper demonstrates the implementation of dataflow analysis in Datafun, a language that combines Datalog and higher-order functional programming (Arntzenius and Krishnaswami 2016). By interweaving Datalog with functional programming, these languages are related to functional logic programming (Antoy and Hanus 2010).

There has also been recent interest in automatically refining the abstractions used in Datalog analyses (Zhang et al. 2014) and in example-based synthesis of analysis implementations in Datalog (Albarghouthi et al. 2017).

### 2.2 Constraint logic programming

Constraint logic programming (CLP) extends logic programming with constraint solving over various theories (Jaffar and Lassez 1987; Jaffar and Maher 1994; Gavanelli and Rossi 2010). Because they both provide mechanisms for reasoning about logical formulae, there are some similarities between FormuLog and CLP. However, FormuLog takes a substantially different approach to constraint solving than traditional CLP, in which constraints are represented by distinguished predicates. When the CLP runtime encounters one of these predicates during evaluation, it adds the relevant constraint to its constraint store and checks that the store is still consistent. This means that a constraint is evaluated in the context of all the constraints that have come before it, and there is no way to manipulate, propagate, or reason about constraints outside of this mechanism. On the other hand, by reifying logical constraints as terms and only treating them as constraints when applied to certain functions, FormuLog provides the programmer with more control over how constraints should be handled. This flexibility is necessary for some static analyses. For instance, model checkers in the style of BLAST (Henzinger et al. 2002) dynamically associate logical formulae with individual program points (that is, code locations in the program under analysis). These analyses must be able to explicitly refer to and manipulate the formula associated with a given program point, and might want to reason about a formula in a local (non-global) context. It would be hard to meet these requirements using a traditional CLP approach in which constraints are represented by predicates and interpreted within a global store.

Some CLP-like systems do represent constraints as terms (Codish et al. 2008; Holzbaur 1995), and some of them have been used to implement static analysis algorithms such as abstraction refinement model checking (Podelski and Rybalchenko 2007). Interestingly, FormuLog represents constraints using \textit{ground} terms, while these systems represent constraints using (in general) \textit{unground} terms, since in these systems constraint-level variables are CLP-level variables. In program analysis, where we want to reason about constraints over variables in the input program, this has the effect of punning logic programming-level variables with input program variables. While
FormuLog: Datalog for static analysis involving logical formulae

define type 'A tree = leaf | node('A tree, 'A, 'A tree).

declare fun sum(i32 tree) : i32.
fun sum(Tree) =
  match Tree with
  | leaf => 0
  | node(L, V, R) => V + sum(L) + sum(R)
end.

declare input num_tree(i32 tree).
num_tree(node(leaf, 42, leaf)).
num_tree(node(node(leaf, 1, leaf), 3, node(leaf, 5, leaf))).

declare output tree_sum(i32 tree, i32).
tree_sum(Tree, Sum) :- num_tree(Tree), sum(Tree) = Sum.

Fig. 2. A FormuLog program consists of type-related metadata and the definitions of functions and relations.

this has some nice benefits, such as being able to use unification to avoid explicitly renaming variables in formulae (Podelski and Rybalchenko 2007), it does require additional bookkeeping by the analysis to ensure that constraints are formed over the correct variables. As this bookkeeping involves propagating information about logic programming-level variables, it seems difficult to use this approach in a Datalog-like setting, where Datalog’s range restriction does not allow a fact that has an unbound variable to be derived (and thus propagated).

Furthermore, since many CLP implementations are extensions of Prolog (Gavanelli and Rossi 2010), they inherit the limitations of Prolog’s depth-first search evaluation strategy. For example, the termination of an analysis written in one these CLP languages is sensitive to the order of clauses and the order of atoms within a clause body, a problem not faced by Datalog. Additionally, since they are not as tightly coupled to a particular evaluation strategy, Datalog-based systems like FormuLog are freer to apply optimizations that can help scale static analyses, such as analysis rewriting (Bravenboer and Smaragdakis 2009) and parallelization (Scholz et al. 2016).

3 Language design

The goal of FormuLog is to make it possible to represent, manipulate, and reason about logical formulae. To achieve this goal, FormuLog extends the grammar of Datalog terms with n-ary constructors and function calls, and imposes a type system to enforce the correct use of these more complex types of terms. Section 3.1 introduces FormuLog’s approach to these features, which are also present, to one extent or another, in previous variants of Datalog for static analysis (Scholz et al. 2016; Madsen et al. 2016). Section 3.2 describes how FormuLog combines these features with built-in support for SMT solving to enable computation involving logical formulae. Section 3.3 concludes with a short discussion of the semantics of FormuLog.

3.1 Types, constructors, functions, and relations

We use a toy program to help introduce FormuLog’s use of types, constructors, functions, and relations (Figure 2). A FormuLog program conceptually consists of two parts: type-related meta-
A FormuLog program represents logical formulae using ground terms, manipulates formulae with user-defined functions, and reasons about formulae through built-in SMT support. We discuss each in turn.

### 3.2.1 Representing logical formulae

FormuLog currently has support for constructing formulae concerning booleans and 32-bit bit vectors, although it would be entirely possible to extend the formula language to support reasoning about floating point numbers, arrays, etc. These formulae are created using the constructors belonging to two built-in algebraic data types (Figure 3). The type \texttt{bool\_exp} represents nor-
mal logic connectives and (signed) comparisons between 32-bit bit vector expressions, which are represented by the type \( \text{bv32}_{\text{exp}} \). The constructor \( \text{bv32}_{\text{const}}(N) \) takes a 32-bit integer as an argument and represents a bit vector constant of value \( N \). The constructor \( \text{bv32}_{\text{sym}}(S) \) represents a symbolic bit vector identified by a string \( S \). For example, the term

\[
\text{and(bv32}_{\text{eq}}(\text{bv32}_{\text{sym}}("x"), \text{bv32}_{\text{const}}(42)), \\
\text{bv32}_{\text{eq}}(\text{bv32}_{\text{sym}}("y"), \text{bv32}_{\text{add}}(\text{bv32}_{\text{sym}}("x"), \text{bv32}_{\text{const}}(1))))
\]

represents the constraint \( x = 42 \land y = x + 1 \), where \( x \) and \( y \) are bit vectors.

### 3.2.2 Manipulating logical formulae

Because formulae are normal terms, they can be manipulated using standard functions. For instance, this function performs a symbol substitution in a bit vector expression:

```latex
\text{declare fun subst(string, string, bv32}_{\text{exp}} : bv32}_{\text{exp}}.\\
\text{fun subst(Old, New, E) =}\\
\text{match E with}\\
\text{ | bv32}_{\text{const}}(N) => E}\\
\text{ | bv32}_{\text{sym}}(S) => if S == Old then bv32}_{\text{sym}}(New) else E}\\
\text{ | bv32}_{\text{neg}}(E1) => bv32}_{\text{neg}}(subst(Old, New, E1))}\\
\text{...}\\
\text{end.}
```

The infix function \( \iff \) returns true if its two arguments reduce to the same term.

### 3.2.3 Reasoning about logical formulae

FormuLog provides built-in functions that support reasoning about logical formulae. The function \( \text{is}_{\text{sat}}(t) \) takes as an argument a term of type \( \text{bool}_{\text{exp}} \) and returns a boolean indicating whether that term is satisfiable. For example, say that the variable \( Z \) is bound to the formula presented above that represents the constraint \( x = 42 \land y = x + 1 \). In this case, \( \text{is}_{\text{sat}}(Z) \) evaluates to true, since there exist values for \( x \) and \( y \) that make the formula true when interpreted under the theory of bit vectors (namely, 42 and 43). On the other hand, the invocation \( \text{is}_{\text{sat}}(\text{and}(Z, \text{false})) \) evaluates to the boolean false. Our current prototype also includes a function for computing Craig interpolants (Craig 1957), which are useful for abstract model checking (McMillan 2006), and we anticipate adding support for additional logical queries, such as finding a model for a formula.

### 3.3 Semantics

Although we have not yet formally established the semantics of FormuLog, we conjecture that the meaning of a well-formed FormuLog program is the least model of the theory that corresponds to the program. This would give it a semantics similar to that of a Datalog program. However, this model could differ from a model corresponding to a Datalog program in two significant respects. First, since FormuLog has \( n \)-ary constructors, the least model might be infinite. Currently it is the programmer’s responsibility to ensure that the model is finite, but we hope to develop analyses that will warn the programmer if it appears that the model might be infinite.
A. Bembenek and S. Chong

define type node = i32.
define type reg = string.
define type cond = cond_eq | cond_ne | cond_lt | cond_le | ...
define type inst = inst_jmp(cond, reg, reg, node) | inst_fail | ...
declare input stmt(node, inst).
declare input fall_thru_succ(node, node).

define type store = (reg, bv32_exp) map.
define type state = (store * bool_exp * i32 * i32).
declare input init_fuel(i32).
declare input start(node, store).

Fig. 4. Program instructions can be represented using algebraic data types and a CFG can be represented using EDB relations. The symbolic executor uses complex terms to keep track of program state; initial state is set by the value of EDB relations.

(while not issuing too many false alarms). Second, since FormuLog has interpreted functions, the model would depend on the meaning of the functions in the program and would not be a Herbrand model. This need not necessarily detract from the declarativeness of FormuLog: as long as every function is pure and evaluates to a ground term, then every function has a mathematical meaning that is independent of its evaluation. For example, the meaning of a pure function does not depend on whether a call-by-value or call-by-name reduction strategy is used. As a well-formed FormuLog program is composed of declarative parts, we would expect the program as a whole to have a declarative semantics. In particular, the meaning of a well-formed program should be independent of the order of rules and the order of atoms within rule bodies.

4 Case study: symbolic execution

This section demonstrates how FormuLog can be used to implement symbolic execution. The encoding is natural and relatively concise (about 200 lines of code), and leverages all of the language features described in the previous section.

A symbolic executor interprets a program in which some values are symbolic, in that they represent not a single runtime value but a set of runtime values [King 1976; Cadar et al. 2008; Cadar and Sen 2013]. Typically, the set associated with a value is initially unconstrained, meaning that it can concretely have any value of the relevant type. However, when the symbolic executor reaches a condition that depends on a symbolic value, it will fork into two processes, one in which the condition is true and one in which the condition is false. In each of the forks, it constrains the symbolic value so that it is consistent with the branch that has been taken. For example, say that a variable \( x \) is associated with the unconstrained symbolic value \( \alpha \). When execution encounters the branch \( x < 42 \), execution will split and explore both branches. However, in the “then” branch \( \alpha \) will be constrained to be less than 42, and in the “else” branch it will be constrained to be greater than or equal to 42. By checking to make sure that the accumulated constraints on a symbolic value are consistent, execution can avoid exploring infeasible paths through the program, which saves analysis time and reduces false positives.

Our FormuLog implementation of symbolic execution takes as input a control flow graph (CFG) representation of a simple register language (Figure 4 top half). The input relation \texttt{stmt} relates a CFG node to an instruction in this language. Although our full implementation supports
declare output reach(node, state).
declare output step_to(node, state).

reach(Node, New_state) :-
    step_to(Node, State),
    decr_fuel(State) = some(New_state).

step_to(Succ, New_state) :-
    reach(Node, State),
    stmt(Node, inst_jmp(Cond, Val1, Val2, Succ)),
    New_state = add_cond_to_state(Cond, Val1, Val2, State),
    is_sat(get_constraints(New_state)) = true.

declare fun add_cond_to_state(cond, reg, reg, state) : state.
fun add_cond_to_state(Cond, Reg1, Reg2, State) =
    let (Store, Constraints, Count, Fuel) = State in
    let some(Val1) = get(Reg1, Store) in
    let some(Val2) = get(Reg2, Store) in
    let Constraint =
        match Cond with
        | cond_eq => bv32_eq(Val1, Val2)
        | cond_ne => not(bv32_eq(Val1, Val2))
        ...
    end in
    (Store, and(Constraint, Constraints), Count, Fuel).

declare output failed_assert(node, state).
failed_assert(Node, State) :-
    reach(Node, State),
    stmt(Node, inst_fail).

Fig. 5. The evaluation of the symbolic executor is defined using a combination of relations and functions.

a wider range of instructions (such as unary and binary arithmetic operations), here we focus
on two instructions: a conditional jump instruction that steps to a node based on the result of
comparing two registers, and a fail instruction, which represents that execution has reached a
failed assertion and a bug has been found. The relation fall thru succ relates a node to its
non-jump successor in the CFG.

The symbolic interpreter needs to maintain state that tracks the current value of each register
and any constraints on symbolic values (Figure 4, bottom half). Accordingly, the type state
consists of a store (a map from registers to bool_exp, implemented as an association list) and a
bool_exp that tracks constraints on the program path. The state also includes two integers. The
first exists for a purely technical reason (to create fresh symbols). The second records the amount
of “fuel” left for execution. Each step in the symbolic execution consumes a unit of fuel. This
guarantees that execution will terminate, but also means that the analysis is in general unsound
and may miss some bugs. The amount of fuel to use is set by the EDB relation init_fuel, and
the EDB relation start sets the entry point of the CFG and the initial store.

Two mutually recursive relations define the symbolic interpreter (Figure 5). The relation reach
records that execution has reached a CFG node with a particular execution state, while the relation \texttt{step-to} records that execution is attempting to take a step to a particular node. The relation \texttt{reach} is defined by a single rule, which says that a node is reachable if execution is trying to step to it and the execution state has enough fuel (the function \texttt{decr.fuel} returns the option constructor \texttt{none} if the state is out of fuel). On the other hand, there is a different \texttt{step-to} rule for each possible execution step. The rule we give here describes performing a conditional jump. The function \texttt{add.cond.to.state} adds a new constraint on the execution path representing the condition for the jump, and the function \texttt{is.sat} is used to make sure that the condition is actually satisfiable given the state’s accumulated constraints. Another rule would be necessary to handle the case when the negation of the condition is satisfiable. Since these cases are not mutually exclusive, both rules can be triggered simultaneously, which corresponds to the symbolic executor forking into multiple processes. Finally, the rule \texttt{failed.assert} is triggered when a fail instruction is reachable.

It was very straightforward to implement the symbolic interpreter presented in this section. Of course, it is for a toy language, and in future work we hope to evaluate whether FormuLog allows a similarly clean encoding of symbolic execution for a more complex language, like JVM bytecode. We also hope to test whether an analysis written in FormuLog can be sufficiently performant. For example, successful symbolic execution engines typically use heuristics to focus on high-priority areas of the execution state space. It would be interesting to see if these heuristics could be encoded in FormuLog and, if not, what sort of language features could be added to FormuLog to make this possible while still keeping the language declarative.

In addition to the symbolic interpreter described in this section, we have implemented an abstract model checker that operates over a similar register language. The model checker is based on the lazy abstraction paradigm \cite{Henzinger:2002} and uses interpolants to refine abstractions \cite{McMillan:2006}. The implementation is about 500 lines of FormuLog.

5 Prototype

We have written a prototype implementation of FormuLog in approximately 6,000 lines of Java code. Here we outline the system, give some initial performance numbers, and propose some potential optimizations.

5.1 System design

The FormuLog prototype works in five stages: parsing, type checking, program validation, program rewriting, and evaluation. FormuLog source code is first parsed into an abstract syntax tree (AST) representation using a generated parser. The AST is then type checked using a unification-based algorithm. This involves type checking the rules, facts, and function definitions against the type signatures declared in the program metadata. All type information is discarded after type checking. Next, a validator checks that the AST represents a well-formed program. In particular, it ensures that the program meets the requirements of stratified negation and some restrictions on the use of variables. In addition to the standard range restriction of Datalog, FormuLog requires that every variable that appears within an argument to a function is bound elsewhere in the clause. This is necessary since functions in FormuLog are not invertible in general. Finally, a rule preprocessor reorders the premises in each rule so that the premises can be evaluated from left to right, ensuring that a variable is bound before being used in a function call. Rewriting rules
Table 1. While not exceptionally fast, the FormuLog prototype achieved respectable parallel scaling on two test cases. This figure shows absolute times, speedup relative to using a single thread, and efficiency (the ratio of speedup to the number of threads).

| Number of threads | Transitive closure | Symbolic execution |
|-------------------|--------------------|--------------------|
|                   | Time (s)           | Speedup            | Efficiency       |
| 1                 | 303                | 1.0                | 1.00             |
| 8                 | 49                 | 6.2                | 0.77             |
| 16                | 27                 | 11.2               | 0.70             |
| 24                | 24                 | 12.6               | 0.53             |
| 32                | 21                 | 14.4               | 0.45             |
| 40                | 19                 | 15.9               | 0.40             |
| 48                | 18                 | 16.8               | 0.35             |

|                   | Time (s)           | Speedup            | Efficiency       |
| 1                 | 90                 | 1.0                | 1.00             |
| 8                 | 17                 | 5.3                | 0.66             |
| 16                | 12                 | 7.5                | 0.47             |
| 24                | 10                 | 9.0                | 0.38             |
| 32                | 10                 | 10.0               | 0.28             |
| 40                | 9                  | 10.0               | 0.25             |
| 48                | 9                  | 10.0               | 0.21             |

in this way is only safe if every function is pure (which, in the context of FormuLog, means that every function must evaluate to a ground term).

The FormuLog interpreter evaluates one stratum of the program at a time. Each stratum is evaluated using a parallelized bottom-up algorithm inspired by pipelined semi-naive evaluation (Loo et al. 2006). Parallelism is effected through a work-stealing thread pool. Each work item is a rule that has been partially evaluated on a fact; a worker thread completes a work item by fully evaluating the rule against a database of currently derived facts. When evaluation results in a fact that has not been seen before, each relevant rule is partially evaluated on that fact and then submitted as a new work item. When a logical function like \( \text{is\_sat} \) is invoked in the course of evaluation, the arguments are translated into Z3 formulae and a query is made to the Z3 SMT solver.\(^2\)

### 5.2 Initial performance results

We have tested the performance and scalability of our prototype on two programs and, while our prototype was not exceptionally fast, it did achieve a respectable level of parallel scaling in both cases. All measurements were taken on 2.5 GHz Intel Xeon Platinum 8175 machine running Ubuntu 16.04, with 192 GiB of memory and 24 hyper-threaded physical cores. We varied the size of the work-stealing thread pool from one to 48 threads by multiples of eight. We performed three trials for each test configuration, and report the median result here.

The first test program computed the transitive closure of a random directed graph (Table [1] top). The program was written in a pure-Datalog fragment of FormuLog and never invoked the SMT solver. The graph had 2,000 vertices, and each potential edge was included in the graph with a probability of 0.1. This very dense graph was strongly connected, so that the transitive closure relation had four million tuples. Performance improved with the size of the thread pool, although the speedup was sublinear. Using 48 threads, FormuLog computed the transitive closure relation in 18 seconds, a speedup of 16.7x relative to using a single thread. For comparison, Soufflé (Scholz et al. 2016) computed the transitive closure relation in about two seconds while achieving a speedup of around 3x (using 48 threads versus a single thread); on the other hand, even with 48 threads Flix (Madsen et al. 2016) did not finish in a reasonable amount of time.

\(^2\) Z3 is available at [https://github.com/Z3Prover/z3](https://github.com/Z3Prover/z3)
A. Bembenek and S. Chong

The second test ran the symbolic execution implementation from Section 4 (Table 1, bottom). It achieved a max speedup of 10x, when it was able to explore over 5,800 program states in about nine seconds. The symbolic interpreter was invoked on an input program with a high branching factor (a tight loop in the program forces the symbolic interpreter to fork into three processes each iteration). Consequently, the symbolic executor had to make a call to Z3 at nearly every step. We conjecture that this test achieved a lower speedup relative to the transitive closure example because of the higher percentage of work done outside of our parallel algorithm, such as in interprocess communication and constraint solving.

5.3 Future optimizations

The performance of our prototype has been acceptable on small toy examples, but we anticipate that making it scale will require some smart optimizations. We have identified five main areas that we would like to focus on: memoization, SMT solving, database optimizations, work stealing, and unification.

5.3.1 Memoization

Our prototype memoizes terms and atoms during creation, so that there are never multiple Java objects that represent terms or atoms that are syntactically the same. The advantage to this approach is that the equality of two terms (or atoms) and the hash code of a term (or atom) can be computed in constant time. Since these operations happen often during rule evaluation, we expect memoization to have a net benefit on performance, outweighing the downside that a shared cache needs to be checked every time a term or atom is created. However, more experimentation is necessary to determine the impact of memoization on real workloads. It would also be possible to extend memoization to other parts of the runtime, such as caching the results of function calls (which is possible because functions are pure and deterministic).

5.3.2 SMT solving

Making calls to an SMT solver can be expensive and applications that depend on SMT solving benefit by limiting the number of calls they make. For instance, the symbolic execution engine KLEE caches SMT query results such that, in certain cases, it is possible to tell if the result of one SMT query implies another without having to make another SMT call (Cadar et al. 2008). Our prototype currently uses Z3 naively.

5.3.3 Database optimizations

Datalog implementations that can handle industry-scale static analyses depend on database-style optimizations, such as optimizing the order that premises in a rule body are evaluated (akin to optimizing the order of database operations) and using indexed data structures to store Datalog facts (Bravenboer and Smaragdakis 2009; Jordan et al. 2016; Scholz et al. 2016). Our FormuLog implementation does not optimize the order in which it evaluates rule premises and, while it does use an indexed data structure to store facts, this data structure has not been extensively optimized.
5.3.4 Work stealing

We have not experimented with the size of the work items we submit to the work-stealing thread pool. However, we anticipate having to tune the granularity of work items to get optimal performance out of our parallel evaluation algorithm.

5.3.5 Unification

We have developed an algorithm to unify two terms that runs in time linear in the number of subterms (ignoring the time necessary to reduce any functions occurring in the terms). This is tricky in the presence of functions, which can only be evaluated after every variable in their arguments is bound. For example, the calls to the function $f$ in this example force the subterms to be unified in a particular order ($a$ and $b$ are constructors):

\[
a(f(X_n), f(X_{n-1}), \ldots, f(X_0), b(0)) = \\
a(b(X_{n+1}), b(X_n), \ldots, b(X_1), b(X_0))
\]

$b(X_0)$ must first be unified with $b(0)$, then $b(X_1)$ can be unified with the result of $f(X_0)$, then $b(X_2)$ can be unified with the result of $f(X_1)$, and so on. A naive algorithm might make a quadratic number of left-to-right passes over the terms. Although our linear-time algorithm should have performance benefits in theory, in practice a naive algorithm might outperform it on typical unification instances because of constant factors. As an alternative approach, complex unification instances (such as the one above) could be rewritten into a collection of simpler unification instances, which in turn could be solved with a simpler algorithm. As unification is such a common operation in FormuLog evaluation, the performance of the unification algorithm could have a substantial impact on overall performance.

6 Conclusion

FormuLog is an extension of Datalog that makes it possible to represent, manipulate, and reason about logical formulae. In turn, this enables one to declaratively implement static analyses such as symbolic execution and abstract model checking. While we anticipate that major engineering work will be required to scale FormuLog to real-world static analysis problems, we nonetheless believe that languages like it have the potential to help address some concrete challenges faced by modern static analyses.

References

Albarghouthi, A., Koutris, P., Naik, M., and Smith, C. 2017. Constraint-based synthesis of Datalog programs. In International Conference on Principles and Practice of Constraint Programming. Springer, 689–706.

Antoy, S. and Hanus, M. 2010. Functional logic programming. Commun. ACM 53, 4 (Apr.), 74–85.

Arntzenius, M. and Krishnaswami, N. R. 2016. Datafun: A functional Datalog. In Proceedings of the 21st ACM SIGPLAN International Conference on Functional Programming. ICFP 2016. ACM, New York, NY, USA, 214–227.

Bravenboer, M. and Smaragdakis, Y. 2009. Strictly declarative specification of sophisticated pointsto analyses. In Proceedings of the 24th ACM SIGPLAN Conference on Object Oriented Programming Systems Languages and Applications. OOPSLA ’09. ACM, New York, NY, USA, 243–262.
CADAR, C., DUNBAR, D., AND ENGLER, D. 2008. KLEE: Unassisted and automatic generation of high-coverage tests for complex systems programs. In Proceedings of the 8th USENIX Conference on Operating Systems Design and Implementation. OSDI’08. USENIX Association, Berkeley, CA, USA, 209–224.

CADAR, C. AND SEN, K. 2013. Symbolic execution for software testing: Three decades later. Commun. ACM 56, 2 (Feb.), 82–90.

CERI, S., GOTTLOB, G., AND TANCA, L. 1989. What you always wanted to know about Datalog (and never dared to ask). IEEE Transactions on Knowledge and Data Engineering 1, 1, 146–166.

CODISH, M., LAGOON, V., AND STUCKEY, P. J. 2008. Logic programming with satisfiability. Theory and Practice of Logic Programming 8, 1, 121–128.

CRAIG, W. 1957. Three uses of the Herbrand-Gentzen theorem in relating model theory and proof theory. The Journal of Symbolic Logic 22, 3, 269–285.

GAVANELLI, M. AND ROSSI, F. 2010. Constraint logic programming. In A 25-Year Perspective on Logic Programming. Springer-Verlag, 64–86.

GREEN, T. J., HUANG, S. S., LOO, B. T., AND ZHOU, W. 2013. Datalog and recursive query processing. Foundations and Trends in Databases 5, 2, 105–195.

HENZINGER, T. A., JHALA, R., MAJUMDAR, R., AND SUTRE, G. 2002. Lazy abstraction. In Proceedings of the 29th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages. POPL ’02. ACM, New York, NY, USA, 58–70.

HOLZBAUR, C. 1995. OFAI clp(Q,R) manual. Tech. Rep. TR-95-09, Austrian Research Institute for Artificial Intelligence, Vienna. Edition 1.3.3.

JAFFAR, J. AND LASSEZ, J.-L. 1987. Constraint logic programming. In Proceedings of the 14th ACM SIGACT-SIGPLAN Symposium on Principles of Programming Languages. POPL ’87. ACM, New York, NY, USA, 111–119.

JAFFAR, J. AND MAHER, M. J. 1994. Constraint logic programming: A survey. The Journal of Logic Programming 19, 503–581.

JORDAN, H., SCHOLZ, B., AND SUBOTIČ, P. 2016. Soufflé: On synthesis of program analyzers. In International Conference on Computer Aided Verification. Springer, 422–430.

KING, J. C. 1976. Symbolic execution and program testing. Commun. ACM 19, 7 (July), 385–394.

KOWALSKI, R. 1979. Algorithm = logic + control. Commun. ACM 22, 7 (July), 424–436.

LIVSHITS, V. B. AND LAM, M. S. 2005. Finding security vulnerabilities in Java applications with static analysis. In USENIX Security Symposium. Vol. 14. 18–18.

LOO, B. T., CONDIE, T., GAROFALAKIS, M., GAY, D. E., HELLERSTEIN, J. M., MANIATIS, P., RAMAKRISHNAN, R., ROSCOE, T., AND STOICA, I. 2006. Declarative networking: Language, execution and optimization. In Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data. SIGMOD ’06. ACM, New York, NY, USA, 97–108.

MADSSEN, M., YEE, M.-H., AND LHOTÁK, O. 2016. From Datalog to Flix: A declarative language for fixed points on lattices. In Proceedings of the 37th ACM SIGPLAN Conference on Programming Language Design and Implementation. PLDI ’16. ACM, New York, NY, USA, 194–208.

MCMILLAN, K. L. 2006. Lazy abstraction with interpolants. In International Conference on Computer Aided Verification. Springer, 123–136.

PODIELSKI, A. AND RYBALCHENKO, A. 2007. ARMC: the logical choice for software model checking with abstraction refinement. Practical Aspects of Declarative Languages, 245–259.

REPS, T. W. 1995. Demand interprocedural program analysis using logic databases. In Applications of Logic Databases. Springer, 163–196.

SCHOLZ, B., JORDAN, H., SUBOTIĆ, P., AND WESTMANN, T. 2016. On fast large-scale program analysis in Datalog. In Proceedings of the 25th International Conference on Compiler Construction. ACM, 196–206.

TOMAN, J. AND GROSSMAN, D. 2017. Taming the static analysis beast. In LIPIcs-Leibniz International Proceedings in Informatics. Vol. 71. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.
Whaley, J., Avots, D., Carbin, M., and Lam, M. S. 2005. Using Datalog with binary decision diagrams for program analysis. In Asian Symposium on Programming Languages and Systems. Springer, 97–118.

Whaley, J. and Lam, M. S. 2004. Cloning-based context-sensitive pointer alias analysis using binary decision diagrams. In Proceedings of the ACM SIGPLAN 2004 Conference on Programming Language Design and Implementation. PLDI ’04. ACM, New York, NY, USA, 131–144.

Zhang, X., Mangal, R., Grigore, R., Naik, M., and Yang, H. 2014. On abstraction refinement for program analyses in Datalog. In Proceedings of the 35th ACM SIGPLAN Conference on Programming Language Design and Implementation. PLDI ’14. ACM, New York, NY, USA, 239–248.