Automatic Verification of Strongly Dynamic Software Systems

N. Dor[1], J. Field[2], D. Gopan[3], T. Lev-Ami[4], A. Loginov[3], R. Manevich[4], G. Ramalingam[2], T. Reps[3], N. Rinetzky[4], M. Sagiv[4], R. Wilhelm[5], E. Yahav[2], and G. Yorsh[4]

[1] IBM Haifa Research Lab; nurit@il.ibm.com
[2] IBM T.J. Watson Research Center; {jfield,eyahav,grama}@us.ibm.com
[3] Univ. of Wisconsin; {alexey,gopan,reps}@cs.wisc.edu
[4] Tel Aviv Univ.; {tla,rumster,maon,msagiv,gretay}@tau.ac.il
[5] Univ. des Saarlandes; wilhelm@cs.uni-sb.de

Abstract. Strongly dynamic software systems are difficult to verify. By *strongly dynamic*, we mean that the actors in such systems change dynamically, that the resources used by such systems are dynamically allocated and deallocated, and that for both sets, no bounds are statically known. In this position paper, we describe the progress we have made in automated verification of strongly dynamic systems using abstract interpretation with three-valued logical structures. We then enumerate a number of challenges that must be tackled in order for such techniques to be widely adopted.

1 The Problem

We will use the term *strongly dynamic* system to refer to software in which the set of actors in the system changes dynamically, where resources are dynamically allocated and deallocated, and where for both sets no bounds are statically known.

Heap allocation of data structures, which is the principal mechanism for creating structured data in modern languages, is the classical manifestation of a strongly dynamic system. It is well known that manipulating heap-allocated data is error-prone, due primarily to the complexity of potential aliasing relationships among pointer-valued data. However, dynamic resource manipulation occurs at many levels in modern software; such dynamic resources may include, e.g., persistent data in databases, language-level threads, operating system resources such as files and sockets, and web sessions.

Formally, the state of strongly dynamic systems may be viewed as a evolving universe of entities over which the program operates. Due to its evolving character, such universes are difficult to reason about, both for programmers and for automated reasoning tools. This in turn makes automated verification for such programs both important and challenging.

While automatic memory allocation and garbage collection in modern programming languages has eased the burden of correctly managing the lifetime of

heap-allocated memory, reasoning about the *states* of an unbounded number of heap-allocated objects and their interrelationships remains a difficult challenge.

Frequently, strongly dynamic systems are encapsulated in abstract data types that restrict direct access to the underlying evolving universe, and hence allow the programmer to reason only about the data type's interface. In such cases, the principal verification challenge is to ensure that the implementation of the data type correctly realizes the desired abstract properties.

However, in the case of scarce high-level system resources, such as processes, sessions, and buffers, programmers do not have the luxury of automatic resource management or abstract data type encapsulation; instead, they must reason directly about resource state *and* resource lifetime. Verifying the correct usage of such resources is therefore particularly challenging.

Finally, concurrency and distribution drastically complicates the problem of program verification, since both the data *and* control structures of the program operate over unbounded universes.

This position paper sketches the state of art in automatic verification of properties of strongly dynamic systems using abstract interpretation [CC79] with *three-valued logical structures*.

1.1 Program Properties

Our focus is on verifying that programs satisfy certain specific (but not fixed) safety and liveness properties, such as those illustrated below, as opposed to establishing complete correctness of a program (with respect to its complete specification). The progress made in selective "property verification" in recent years makes us cautiously optimistic about its long-term prospects. It poses several challenging research problems, but promises to play an important and relevant role in industrial software-development practice within a reasonable time frame. We are in general interested in the following safety properties:

Memory Cleanness In this case, we wish to prove that a program does not perform pointer manipulations that have unpredictable effects. We call these *cleanness* properties, since they are generic to a given programming model, rather than application-specific (although frequently, in order to show cleanness, it is necessary to prove stronger properties as well). For sequential C-like programs, such properties include: (i) absence of null dereferences, (ii) absence of memory leaks, and (iii) absence of double deallocations. The failure of these properties is frequently exploited by hackers, see, e.g., [Rei03].

The use of garbage collection in modern languages eliminates some of the problems that occur when programmers manage memory allocation and deallocation manually. However, garbage collection does not eliminate the possibility of premature resource depletion due to delayed deallocation.

Similar resource-management problems are possible with other kinds of resources; e.g., database connections, buffers, files, and sockets. In our experience, many serious problems in large applications arise from resources that are freed too late.

Establishing Data-Structure Invariants Data structures built using pointers can be characterized by invariants describing their "shape" at stable states, i.e., between operations defined by their external interfaces. These invariants are usually not preserved by the execution of individual program statements, and it is challenging to prove that data-structure invariants are reestablished after a sequence of statements executes [Hoa75].

Conformance of Library Specifications In cases where a library's interface is accompanied by a formal specification of key assumptions and guarantees, it is useful to statically verify that a particular client satisfies, or *conforms* to the interface properties. One can then choose to verify that a library's implementation satisfies its interface specification (thus enabling modular reasoning and analysis of full systems), or simply treat the interface specification as presumptively correct (thus limiting the scope of verification to the client). While significant progress has been made in client-component conformance verification (e.g., see [CDH+00,DM01,FTA02,FLL+02,AE02,RWF+02,FGRY03,DLS02]), doing precise verification that can scale to large and complex programs is challenging.

Concurrency Concurrent programs introduce a number of challenging verification issues, particularly when the number of concurrent threads may be unbounded. In this context, data and control are strongly related: thread-scheduling information may require an understanding of the structure of the heap (e.g., the structure of the scheduling queue). Also, heap analysis requires information about thread scheduling, because multiple threads may be manipulating the heap simultaneously. In addition to verifying the absence of "generic" concurrency anomalies, such as races and deadlocks, one often wishes to prove application-specific properties of concurrent protocols that are required to hold under arbitrary thread interleavings.

2 What Has Been Achieved So Far

This section summarizes the progress our group has made on property verification using abstract interpretation with three-valued logical structures [SRW02] and the TVLA system [LAS00], a general-purpose abstract-interpretation engine based on three-valued logic.

Abstract interpretation can be used for verification by generating an over-approximation to the set of states that can arise in any valid program execution; the property of interest is established if the over-approximation demonstrates that no undesirable state can be reached. Typically, problems are cast as a set of equations over a semi-lattice of program properties, and solved by means of successive approximation, possibly with extrapolation.

In [SRW02], we showed that first-order logic can be viewed as a parametric framework for defining both the semantics of a program and for expressing a variety of properties to be verified. In this framework, concrete program states are represented by logical structures. Three-valued logic, which adds an "unknown"

value to the Boolean values of ordinary two-valued logic, is a natural framework for defining sound, finitary *abstractions* of two-valued structures for the purpose of abstract interpretation.

Memory Cleanness The first application of TVLA was to show memory cleanness of C programs [DRS00]. The algorithm is rather precise in the sense that it yields very few false alarms but it was only applied to small programs.

Interprocedural Analysis [RS01] handles procedures by explicitly representing stacks of activation records as linked lists, allowing rather precise analysis of recursive procedures. However, it does not scale very well. [JLRS04] handles procedures by summarizing their behavior. [RBR+05] presents a new concrete semantics for programs that manipulate heap-allocated storage which only passes "local" heaps to procedures. A simplified version of this semantics is used in [RSY05] to perform more modular summarization by only representing reachable parts of the heap.

Concurrent Java Programs [Yah01] presents a general framework for proving safety properties of concurrent Java programs with an unbounded number of objects and threads. In [YS03] this approach is applied to verify partial correctness of concurrent-queue implementations.

Temporal Properties [YRSW03] proposes a general framework for proving temporal properties of programs by representing program traces as logical structures. A more efficient technique for proving local temporal properties is presented in [SYKS03] and applied to compile-time garbage collection in JavaCard programs.

Correctness of Sorting Implementations In [LARSW00], TVLA is applied to analyze programs that sort linked lists. It is shown that the analysis is precise enough to discover that (correct versions) of bubble-sort and insertion-sort procedures do, in fact, produce correctly sorted lists as outputs, and that the invariant "is-sorted" is maintained by list-manipulation operations such as merge. In addition, it is shown that when the analysis is applied to erroneous versions of bubble-sort and insertion-sort procedures, it is able to discover the error. In [LRS05], abstraction refinement is used to *automatically* derive abstractions that are successfully used to prove partial correctness of several sorting algorithms. The derived abstractions are also used to prove that the algorithms possess additional properties, such as stability and anti-stability.

Conformance to API Specifications [RWF+02] shows how to verify that client programs using a library conform to the library's API specifications. In particular, an analysis is provided for verifying the absence of concurrent-modification exceptions in Java programs that use Java collections and iterators. In [YR04], separation and heterogeneous abstraction are used to scale the verification algorithms and to allow verification of larger programs (several thousands lines of code) that use libraries such as JDBC.

Computing Intersections of Abstractions [Arn04] considers the problem of computing the intersection (meet) of heap abstractions, namely the greatest lower bound of two sets of 3-valued structures. This problem proves to have many applications in program analysis such as interpreting program conditions, refining abstract configurations, reasoning about procedures [JLRS04], and proving temporal properties of heap-manipulating programs, either via greatest-fixed-point approximation over trace semantics or in a staged manner over the collecting semantics. [Arn04] describes a constructive formulation of meet that is based on finding certain relations between abstract heap objects. The enumeration of those relations is reduced to finding constrained matchings over bipartite graphs.

Efficient Heap Abstractions and Representations [MRF+02] addresses the problem of space consumption in first-order state representations by describing and evaluating two new representation techniques for logical structures. One technique uses ordered binary decision diagrams (OBDDs); the other uses a variant of a functional map data structure. The results show that both the OBDD and functional implementations reduce space consumption in TVLA by a factor of 4 to 10 relative to the original TVLA state representation, without compromising analysis time.

[MSRF04] presents a new heap abstraction that works by merging shape descriptors according to a partial isomorphism similarity criterion, resulting in a partially disjunctive abstraction. This abstraction provides superior performance compared to the powerset heap abstraction, without any loss of precision, for a suite of TVLA benchmark verification problems.

[MYRS05] provides a family of simple abstractions for potentially cyclic linked lists. In particular, it provides a relatively efficient predicate abstraction that allows verification of programs that manipulate potentially cyclic linked lists.

Abstracting Numerical Values [GDN+04] presents a generic solution for combining abstractions of numeric and heap-allocated storage. This solution has been integrated into a version of TVLA. In [GRS05], a new abstraction of numeric values is presented, which like canonical abstraction tracks correlations between aggregates and not just indices. For example, it can identify loops that perform array-kills (i.e., assign values to a an entire array). This approach has been generalized to define a family of abstractions (for relations as well as numeric quantities) that is more precise than pure canonical abstraction and allows the basic idea from [GDN+04] to be applied more widely [JGR05].

Assume-Guarantee Reasoning One of the potential ways to scale up shape analysis is by applying it to smaller pieces of code using specifications. [YRS04] presents a new algorithm that takes as input a shape descriptor (describing some set of concrete stores X) and a precondition p, and computes the most-precise shape descriptor for the stores in X that satisfy p. This combines abstract interpretation and theorem provers in a novel way. A prototype has been implemented in TVLA, using the SPASS theorem prover.

Safety Properties of Mobile Ambients The mobile ambient calculus was introduced in [CG98]. In [NNS00], TVLA was applied to prove safety properties programs in the ambient calculus. The main idea is to code the ambient calculus using two-valued logic, and then use TVLA to obtain a sound over-approximation by reinterpreting the logical formulas in Kleene's three-valued logic.

3 Some Remaining Challenges

We have found the framework of abstract interpretation based on three-valued logic to be remarkably powerful, both in its ability to provide a natural formal framework for reasoning about dynamic resource manipulation, and as a substrate for developing efficient data structures and algorithms for verification of such properties. We believe that the formal and practical strengths of this framework should provide a strong base for further research in verification of strongly dynamic systems in the future. In this section, we outline some of the remaining research challenges in this framework.

Scalability and Precision For most of the properties discussed in prior sections, automatic verification of a software system of significant size (e.g., web servers, operating systems, or a *compiler*) remains infeasible. The main problem is the scalability of the existing techniques. We believe that we are likely to make steady advances in the scalability of our techniques by (1) exploiting locality in abstractions (e.g., for interprocedural analysis), (2) exploiting compositionality, i.e., exploiting proven properties of small components or ADTs in verification of large systems, (3) dealing with state explosion caused by interleaving of concurrent threads, and (4) developing improved algorithms for manipulating first-order structures.

Determining the properties that are relevant to the verification problem and identifying the objects that need to be reasoned about at any given program point is key to scalable verification using abstract interpretation. This fundamental problem of "choosing the right set of abstractions" appears to be shared by other verification techniques (including deductive approaches) as well. We believe that machine-learning techniques provide one promising *automated* approach to this problem [LRS05]. We are also investigating the use of counterexample-guided abstraction refinement [CGJ+00] to address this problem in an automated fashion.

Another approach to effective abstraction selection is to induce programmers to annotate programs with information about properties or abstractions relevant to the problem at hand. Currently, programmers have little incentive to add annotations defining properties or abstractions of interest, since the benefit of doing so using current verification technology is low. However, as the power of verification techniques to perform state-space exploration begins to scale to programs of realistic size, a cycle of positive reinforcement will arise: programmers will be encouraged to annotate their program with properties of utility to verifiers, because by doing so they will receive accurate and precise feedback

on critical aspects of program correctness, which will in turn make them more productive programmers. Strong type systems provide a precedent: while programmers were initially skeptical of the benefits of strong typing, there is now little disagreement over its value.

Usability Even in cases where automated verification is sufficiently scalable, there are a number of usability challenges: (1) For automatic verification to become an accepted part of everyday programming, it must provide useful feedback as quickly as current compilers generate type errors. (2) In cases where verification fails, counterexamples and error explanations must guide programmers quickly to potential sources of errors. (3) Particularly in safety-critical systems, the trusted code base used by a verifier must itself be verified.

Hybrid Verification Techniques Theorem proving techniques, e.g., [FLL+02], have proved extremely useful for verifying properties of programs equipped with user-specified annotations (e.g., procedure pre- and post-conditions, and loop invariants). The power of such techniques derives from their ability to reason precisely about large collections of program states using symbolic techniques. However, such approaches are less successful in the absence of annotations, particularly when induction is required. Some initial steps have been taken at combining theorem proving and abstract interpretation (or model checking), e.g., [RWF+02]; further work aimed at exploiting the complementary strengths of these approaches seems desirable.

Combined Static and Dynamic Analysis Although dynamic analysis (i.e., instrumentation of code execution to detect anomalies at runtime) cannot by itself prove a program correct, the results of dynamic analysis could be used to suggest certain properties, e.g., loop or class invariants, which would then be statically verified as a component of a larger verification problem.

Restricted Specification Formalisms While it is tempting to allow program properties of interest to be specified using highly expressive formalisms such as first-order logic, by focusing on a more limited set of properties (e.g., *typestate* properties [SY86]), it is possible that more precise and scalable verification techniques could be developed for this class of properties than would be possible in the more general setting.

Other Issues Some of the other challenges in making verification tools useful include: (1) The need to deal with missing source code (e.g., proprietary libraries) (2) Analyzing open programs and modeling the environment (3) Verifying distributed applications.

References

[AE02] K. Ashcraft and D. Engler. Using programmer-written compiler extensions to catch security holes. In *Proc. IEEE Symp. on Security and Privacy*, Oakland, CA, May 2002.

[Arn04] G. Arnold. Combining heap analyses by intersecting abstractions. Master's thesis, Tel Aviv University, October 2004.

[CC79] P. Cousot and R. Cousot. Systematic design of program analysis frameworks. In *Proc. Symp. on Principles of Prog. Languages*, pages 269–282, New York, NY, 1979. ACM Press.

[CDH+00] J. Corbett, M. Dwyer, J. Hatcliff, C. Pasareanu, R., S. Laubach, and H. Zheng. Bandera: Extracting finite-state models from Java source code. In *Proc. Intl. Conf. on Software Eng.*, pages 439–448, June 2000.

[CG98] L. Cardelli and A.D. Gordon. Mobile ambients. In M. Nivat, editor, *Proc. of Foundations of Software Science and Computation Structures (FoSSaCS)*, volume 1378 of *LNCS*, pages 140–155. Springer-Verlag, March 1998.

[CGJ+00] E.M. Clarke, O. Grumberg, S. Jha, Y. Lu, and H. Veith. Counterexample-guided abstraction refinement. In *Proc. Computer Aided Verification*, pages 154–169, 2000.

[DLS02] M. Das, S. Lerner, and M. Seigle. ESP: Path-sensitive program verification in polynomial time. In *Proc. Conf. on Prog. Lang. Design and Impl.*, pages 57–68, June 2002.

[DM01] R. DeLine and M.Fähndrich. Enforcing high-level protocols in low-level software. In *Proc. Conf. on Prog. Lang. Design and Impl.*, pages 59–69, June 2001.

[DRS00] N. Dor, M. Rodeh, and M. Sagiv. Checking cleanness in linked lists. In *Proc. Static Analysis Symp.* Springer-Verlag, 2000.

[FGRY03] J. Field, D. Goyal, G. Ramalingam, and E. Yahav. Typestate verification: Abstraction techniques and complexity results. In *Proc. of SAS'03*, volume 2694 of *LNCS*, pages 439–462. Springer, June 2003.

[FLL+02] C. Flanagan, K.R.M. Leino, M. Lillibridge, G. Nelson, J.B. Saxe, and R. Stata. Extended static checking for java. In *Proc. Conf. on Prog. Lang. Design and Impl.*, pages 234–245, Berlin, June 2002.

[FTA02] J.S. Foster, T. Terauchi, and A. Aiken. Flow-sensitive type qualifiers. In *Proc. Conf. on Prog. Lang. Design and Impl.*, pages 1–12, Berlin, June 2002.

[GDN+04] D. Gopan, F. DiMaio, N.Dor, T. Reps, and M. Sagiv. Numeric domains with summarized dimensions. In *Tools and Algs. for the Construct. and Anal. of Syst.*, pages 512–529, 2004.

[GRS05] D. Gopan, T. Reps, and M. Sagiv. Numeric analysis of array operations. In *Proc. Symp. on Principles of Prog. Languages*, 2005.

[Hoa75] C.A.R. Hoare. Recursive data structures. *Int. J. of Comp. and Inf. Sci.*, 4(2):105–132, 1975.

[JGR05] B. Jeannet, D. Gopan, and T. Reps. A relational abstraction for functions. In *To appear in Proc. 12th Int. Static Analysis Symp*, September 2005.

[JLRS04] B. Jeannet, A. Loginov, T. Reps, and M. Sagiv. A relational approach to interprocedural shape analysis. In *Proc. Static Analysis Symp.* Springer, 2004.

[LARSW00] T. Lev-Ami, T. Reps, M. Sagiv, and R. Wilhelm. Putting static analysis to work for verification: A case study. In *Int. Symp. on Softw. Testing and Analysis*, pages 26–38, 2000.

[LAS00] T. Lev-Ami and M. Sagiv. TVLA: A system for implementing static analyses. In *Proc. Static Analysis Symp.*, pages 280–301, 2000.

[LRS05] A. Loginov, T. Reps, and M. Sagiv. Learning abstractions for verifying data-structure properties. In *Int. Conf. on Computer Aided Verif.*, 2005.

[MRF+02] R. Manevich, G. Ramalingam, J. Field, D. Goyal, and M. Sagiv. Compactly representing first-order structures for static analysis. In *Proc. Static Analysis Symp.*, pages 196–212, 2002.

[MSRF04] R. Manevich, M. Sagiv, G. Ramalingam, and J. Field. Partially disjunctive heap abstraction. In *Proceedings of the 11th International Symposium, SAS 2004*, volume 3148 of *Lecture Notes in Computer Science*, pages 265–279. Springer, aug 2004.

[MYRS05] R. Manevich, E. Yahav, G. Ramalingam, and M. Sagiv. Predicate abstraction and canonical abstraction for singly-linked lists. In *Proceedings of the 6th International Conference on Verification, Model Checking and Abstract Interpretation, VMCAI 2005*, Lecture Notes in Computer Science. Springer, jan 2005.

[NNS00] F. Nielson, H.R. Nielson, and M. Sagiv. A Kleene analysis of mobile ambients. In G. Smolka, editor, *Proc. of ESOP 2000*, volume 1782 of *LNCS*, pages 305–319. Springer, 2000.

[RBR+05] N. Rinetzky, J. Bauer, T. Reps, M. Sagiv, and R. Wilhelm. A semantics for procedure local heaps and its abstractions. In *Proc. Symp. on Principles of Prog. Languages*, 2005.

[Rei03] F. Reig. Detecting security vulnerabilities in C code with type checking (extended abstract). Available at "http://www.cs.nott.ac.uk/∼fxr/", 2003.

[RS01] N. Rinetskey and M. Sagiv. Interprocedural shape analysis for recursive programs. In R. Wilhelm, editor, *Proc. Intl. Conf. on Compiler Construction*, volume 2027 of *LNCS*, pages 133–149. Springer-Verlag, 2001.

[RSY05] N. Rinetzky, M. Sagiv, and E. Yahav. Interprocedural shape analysis for cutpoint-free programs. In *SAS'05, Static Analysis Symposium*, 2005.

[RWF+02] G. Ramalingam, A. Warshavsky, J. Field, D. Goyal, and M. Sagiv. Deriving specialized program analyses for certifying component-client conformance. In *Proc. Conf. on Prog. Lang. Design and Impl.*, pages 83–94, 2002.

[SRW02] M. Sagiv, T. Reps, and R. Wilhelm. Parametric shape analysis via 3-valued logic. *ACM Transactions on Programming Languages and Systems*, 24(3):217–298, 2002.

[SY86] R.E. Strom and S. Yemini. Typestate: A programming language concept for enhancing software reliability. *IEEE Trans. Software Eng.*, 12(1):157–171, 1986.

[SYKS03] R. Shaham, E. Yahav, E.K. Kolodner, and Mooly Sagiv. Establishing local temporal heap safety properties with applications to compile-time memory management. In *Proc. of SAS'03*, volume 2694 of *LNCS*, pages 483–503, June 2003.

[Yah01] E. Yahav. Verifying safety properties of concurrent Java programs using 3-valued logic. In *Proc. Symp. on Principles of Prog. Languages*, pages 27–40, 2001.

[YR04] E. Yahav and G. Ramalingam. Verifying safety properties using separation and heterogeneous abstractions. In *Proceedings of the ACM SIGPLAN 2004 conference on Programming language design and implementation*, pages 25–34. ACM Press, 2004.

[YRS04] G. Yorsh, T. Reps, and M. Sagiv. Symbolically computing most-precise abstract operations for shape analysis. In *Tools and Algs. for the Construct. and Anal. of Syst.*, pages 530–545, 2004.

[YRSW03] E. Yahav, T. Reps, M. Sagiv, and R. Wilhelm. Verifying temporal heap properties specified via evolution logic. In *Proc. of the 12th European Symposium on Programming, ESOP 2003*, volume 2618 of *LNCS*, April 2003.

[YS03] E. Yahav and M. Sagiv. Automatically verifying concurrent queue algorithms. In *Workshop on Software Model Checking*, 2003.