Verify Your Runs

Klaus Havelund and Allen Goldberg

Kestrel Technology, Palo Alto, California, USA
http://www.kestreltechnology.com

1 Introduction

A program verifier determines whether a program satisfies a specification. Ideally verification is achieved by static analysis without executing the code. However, program verification is unsolvable in general. The interactive approach, for example with a human guiding a theorem prover, does not scale to current software systems and libraries. Some restricted kinds of specifications can, however, be checked automatically, for example, type definitions. Also static analysis of properties such as un-initialized variables, null-pointer de-referencing, and array-bound violations scales to production programs on the order of hundreds of thousands of lines of code. Even concurrency-related problems such as data races and deadlocks can to some extent be checked statically, although often resulting in false positives. However, going beyond these simple properties to arbitrarily complex behavior specification and scaling to ever-growing production program size is undoubtedly a challenge, and we cannot expect it to be fully achieved within the 15 year time horizon of the challenge.

Hence, we will probably have to accept that parts of the verification task will remain as proof obligations. It is reasonable to not throw such proof obligations away, but to monitor them during program testing, or in the operations phase. In the latter case, one can program reactions to property violations to achieve some form of fault protection. We call the scientific discipline that studies monitoring properties during program execution *runtime verification* [1]. Much work has been done in this area within recent years.

In this paper we shall outline and classify some current approaches to runtime verification and describe our contributions. We shall describe how we intend to further contribute to this work in the framework of the Grand Verification Challenge. The paper does not address the topic of test *case generation* although runtime verification is a part of this subject. That is, an effectful test case generation framework needs to support the generation of test cases, where a test case consists of inputs to the program together with an oracle that will inspect the program when executed on that input. Generation of the oracle is the runtime verification part. We believe that runtime verification is a rich subject on its own.

2 Specification-Based Runtime Verification

Specification-based runtime verification monitors a program's execution against a *user-provided specification* of intended program behavior. The many approaches to program specification lead to differing styles of runtime verification. One can consider a spectrum of monitoring approaches, ranging from predicate assertions stating properties about a single state at a single program location, to temporal assertions stating properties about temporally separated states at multiple program locations identified by automated program instrumentation. We shall discuss the techniques along four dimensions:

- *Location quantification:* whether the logic allows to quantify over locations in the program to be monitored. Monitors evaluate when certain program locations are reached during program execution. If the monitoring code is executing *in-line*, these locations will contain the monitoring code itself. If monitoring is *off-line*, the locations will contain event generators, that will send events to the monitors that run in some more or less loose form of synchronization with the code. Locations can either be specified individually, by identifying each of them explicitly, or they can be *quantified* over, as in Aspect Oriented Programming, covering many locations with one declaration, for example "before every call of a method defined in class C, evaluate monitor M".
- *Temporal quantification:* whether the logic allows quantification over time points. For example, one can express properties of the form: "whenever a call of method **close** occurs then in the past there has been a call of the **open** method". Some temporal logics only allow to state ordering relationships, while others go further and allow to state relative or absolute time values.
- *Data quantification:* whether the logic allows binding and referral (forward or backward in time) to values across states. For example, "whenever a call of method **close**(f) occurs, with a file argument f, then in the past there has been a call of **open**(f) of the same file f." Obviously, data quantification presumes temporal quantification.
- *Abstract data specification:* whether abstract states mapping variable names to values can be defined together with an abstraction that relates concrete program states to abstract specification states.

In the following we shall classify a collection of monitoring approaches along these dimensions.

Assertions Runtime checks are assertions inserted at specific locations in the code. Assertions were introduced in Java 1.4 such that the programmer can write assertions of the form **assert** ψ for a Java predicate ψ, at explicit program locations. Assertions do not directly support location, temporal or data quantification, nor abstract data specification.

Pre-Post Conditions Pre-post conditions is an extension of the assert statement, where the programmer explicitly indicates where checks should be performed, namely before and after method calls, hence not supporting location quantification. However, a post condition typically relates the value of variables at the start and the end of the method. Thus this is a restricted form of temporal and data quantification. The Eiffel language [23] has long embodied this idea, and recently so has JML [39], the annotation language for Java. In Eiffel there is no provision for abstract data specification, while in JML there is. The Larch Shared Language approach [41] supports abstract (axiomatic) data specification in combination with pre-post conditions.

Invariants Invariants, as found in for example Eiffel and JML, express properties about a single state, and are required to hold at *all* locations where data consistency can be expected, specifically at the completion of method calls, and at the limit after every variable update. Hence this is an example of a logic supporting location quantification. Since an invariant asserts a property of just the current state it does not support temporal or data quantification.

State Machine Notations and Process Algebras In state machines/automata and process algebras, the specification is an *abstract program*, and runtime verification dynamically checks that the executing program is a *refinement* of the abstract program. That is, the instrumented program locations correspond to abstract program states, and monitoring checks the required state sequencing and that each concrete program state satisfies (via an abstraction function) the properties of a corresponding abstract state. This form of specification supports temporal quantification. The Jass system [38] monitors a combination of JML and CSP process algebra. Alternating automata, supporting AND as well as OR states, have shown to be particularly convenient for monitoring logics as demonstrated in [20, 26, 25]. Also state charts [30] offer this combination of AND and OR states. An example illustrating the use of state machines for monitoring is the TLChart system [22], that monitors a combination of temporal logic and state machines. An extension of simple state machines are timed automata, where time constraints can be put on states (one can only be in a state for a certain time period) and on transitions. One such system for runtime verification of timed automata is T-UPPAAL [52] and another similar system is described in [12]. T-UPPAAL also generates test cases. All the systems mentioned above monitor finite traces against finite trace automata. In [18] is described a technique for monitoring against Omega automata: automata that normally accept infinite traces. This is specifically useful for monitoring automata generated from specifications originally targeted for model checkers such as SPIN [51]. In [28] is described an algorithm for synthesizing finite trace monitoring algorithms from LTL specifications, inspired by similar algorithms used for synthesizing infinite trace Omega automata from LTL specifications.

Temporal Assertions While automata and process algebras are operational in nature, temporal logics are declarative. *Temporal logics* have operators that re-

late arbitrary states, and hence support full temporal quantification, and can in many cases allow more succinct specifications. Pre-post conditions support a simple form of temporal quantification by relating two states (the pre-state and the post-state). In the commercial Temporal Rover system [20], one writes past time and future time temporal logic formulas at specific program locations, that get evaluated whenever that program location is reached. This tool hence supports temporal quantification but not location quantification. The MaC system [40] supports temporal past time assertions and location quantification by allowing instrumentation of method calls and variable updates. MaC also allows abstract data specifications referenced as the propositions of the temporal logic. An interesting logic is the future time temporal logic PSL [43] adopted by the hardware industry. In metric temporal logics one can state properties about time. Several such systems have been developed, for example [20, 53]. Regular expressions, and extended regular expressions allowing negation, appear to be very useful for writing certain properties that in temporal logic would become more complicated to state. Such a system is described in [48]. A generalization of metric logics are data logics supporting data quantification, where one can reason about data values existing at different time points. Such systems are described in [21, 25, 19]. Temporal logics are often mapped to automata, although other interpretations are possible, such as for example described in [46, 34], where rewriting is used to interpret temporal logic for monitoring.

General Purpose Specification Languages A monitoring language may be a complete formal specification language, in the style of ASML [6], Maude [15], PVS [44], VDM [55], RAISE [45] or Specware [50]. This is the approach taken at Microsoft where ASML (Abstract State Machine Language) [6] is used for runtime verification as part of a general test case generation framework. Clearly such an approach supports abstract data specification. These full specification languages usually have executable subsets which resemble a programming language, be it functional or state-based. This observation can be exploited by having the specification language be an extension of the programming language, an approach taken in Spec# [8], Microsoft's extension of the work in [6].

3 Predictive Runtime Verification

As with testing, the effectiveness of runtime verification depends on the choice of test suite. For concurrent systems this becomes even more serious because this is compounded by the many possible execution paths of a non-deterministic program. This raises the question of whether there are properties that can be checked on one or a small number of execution traces and still identify bugs with high probability (if such exist). The answer is affirmative due to recent work on what we call *predictive runtime verification*.

In predictive runtime verification a property P to be monitored is replaced with a stronger property Q, i.e. for all inputs x, $Q(x) \rightarrow P(x)$. Furthermore if $\forall x P(x)$ then $Q(x)$ for most x (few false positives) but if $\exists x \neg P(x)$ then $\neg Q(x)$

for most x (good detection). It turns out for certain problems finding such Q is possible.

One of the earliest successes was the Eraser algorithm [47] for detecting data races, that was implemented in Compaq's Visual Threads tool [31]. This algorithm checks a single execution trace in order to determine whether there are any *potentials* for data races: the situation where two threads access a shared variable simultaneously. This work has later been extended to cover other forms of data races, such as higher level data races [4] and atomicity violations [5, 27, 57]. Also deadlocks of the dining philosopher format can be checked in this manner [13]. A generalized predictive analysis framework is presented in [49]. In most of the above mentioned systems, the properties are programmed directly as algorithms in a traditional programming language. Attempts have, however, been made to express the properties in logic [9]. These are often data oriented properties that are best expressed in a monitoring logic appropriate for expressing data quantification and location quantification.

4 Instrumentation

Instrumentation is the modification of the target system with additional code that informs the monitor of events and data values relevant to the monitored properties, such as the taking of a lock, the entry into a method, or the update of a variable. This can be achieved through source code instrumentation, for example using Aspect Oriented Programming as supported by AspectJ [7]; through byte-code instrumentation, BCEL [11] being an example byte-code instrumentation tool; or through object code instrumentation, with Valgrind [54] being an example. The Java-MOP system described in [14] is a generalized framework for instrumenting Java programs specifically for runtime verification. Instrumentation can, however, also be done through debugging interfaces, modification of the runtime system or virtual machine, or through operating system or middleware services.

Naive instrumentation can cause significant degradation of performance and is a significant concern for concurrent and real time systems. In our work we have used byte-code instrumentation and Aspect Oriented Programming to instrument code. Concurrent target systems may be modified by inclusion of wait statements or modifications to schedulers, so that a full range of non-deterministic behaviors are exhibited during testing. This is discussed in the overview paper [24].

5 Our Previous Work

In this section we briefly outline our own and close colleague's work in runtime verification. Some of our early work [32] was done in predictive runtime verification and resulted in a tool for performing predictive deadlock and Eraser-like data race analysis on Java programs, guiding the Java PathFinder (JPF) model checker [56] to confirm the warnings discovered by the much faster predicative

analysis. Instrumentation was done by modifying the Java Virtual Machine of JPF. The work on predictive runtime verification was later re-implemented and elaborated in the Java PathExplorer (Jpax) tool [35, 33]. Specifically the deadlock analysis algorithm was improved to yield fewer false warnings [13]. More recent work on predictive runtime verification includes Cyrille Artho's work that goes beyond low-level data races on single variables, and includes detection of high-level data races on collections of variables [4], and detection of out-dated copies of shared variables [5]. Jpax also supports specification-based runtime verification. The Maude rewriting system [15] is used to define new logics [46, 34]. This has proved extremely elegant since Maude is well suited for defining the syntax and semantics of a logic. In [36, 37] we describe how to synthesize very efficient algorithms based on dynamic programming for monitoring past time logic.

In more recent work we decided to develop a runtime verification framework for Java *in Java*. Eagle [10] is a powerful temporal kernel language supporting temporal quantification and capable of modeling all of the temporal logics and most of the specification paradigms mentioned in this overview. Eagle is an extension of propositional logic with three temporal kernel operators, recursion, and parameterization over formulas in the logic as well as over data values. Formula parameterization allows the user to define new temporal combinators, and hence new temporal logics. The language therefore directly supports the definition of new specification patterns of the kind illustrated in [42]. Data parameterization allows to define properties relating data values from different points in time, hence supporting data quantification. Due to these constructs Eagle can define various forms of past and future time linear temporal logics, real-time logics, interval logics, extended regular expressions and state machines. Eagle furthermore supports abstract data specification in that formulas are interpreted on an abstract state defined as a Java class, and referred to as the Eagle state. In principle there is a stratification of the propositional language and the logic proper so that Java may be replaced by a high-level specification language. The user must define an abstraction mapping from concrete program states to abstract Eagle states. At each instrumentation location in the monitored target system, a method representing the abstraction function is called to update the Eagle state. Noting that Eagle supports the definition of state machines, we see that Eagle hence supports both data refinement and control refinement. A recent extension of Eagle supports automated program instrumentation [17], hence location quantification, using the Aspect Oriented Programming tool AspectJ [7]. In previous work we developed the jSpy tool [29], which instruments Java byte-code. A jSpy instrumentation specification consists of a set of rules, each of which consists of a condition on byte-code and an instrumentation action stating what to report when byte-codes satisfying the condition are executed. The reported events are then picked up by the monitors that in turn check for various user provided properties. Eagle has been used within a test-case framework as described in [2, 3].

6 Future Work

As a scientific discipline specification-based runtime verification does not face the same difficult problems as, say, model checking or theorem proving, and is likely closer to become part of practical software development environments. However, the discipline faces unsolved problems concerned with choice of specification notations, monitoring algorithms, code instrumentation, as well as social issues such the usual resistance amongst software developers to write formal specifications in addition to the code itself. We feel that predictive runtime verification should be part of any development system since it is very effective, fully automated, requires no specifications, and essentially imposes only minor cost to the programmer. The challenge is to identify other problems that lend themselves to this form of analysis. Concerning specification-based runtime verification, choosing the right specification formalism is critical to the success of the approach. The formalism must be simple, yet powerful, and/or, it could be an already accepted notation, such as UML. We will continue experimenting with Eagle, but we will also investigate other formalisms in order to achieve the optimal balance between simplicity, efficiency and effectiveness. Amongst work not mentioned is that on generating specifications from runs [16]. We intend to extend our work in this direction.

References

1. *1st, 2nd, 3rd, 4th and 5th Workshops on Runtime Verification (RV'01 - RV'05), 2001–2005*, volume 55(2), 70(4), 89(2), 113, TBP. of *ENTCS*. Elsevier Science Direct. `http://react.cs.uni-sb.de/rv2005`.
2. C. Artho, H. Barringer, A. Goldberg, K. Havelund, S. Khurshid, M. Lowry, C. Pasareanu, G. Rosu, K. Sen, W. Visser, and R. Washington. Combining Test-Case Generation and Runtime Verification. *Theoretical Computer Science*, 336(2–3):209–234, May 2005. Extended version of [3].
3. C. Artho, D. Drusinsky, A. Goldberg, K. Havelund, M. Lowry, C. Pasareanu, G. Roşu, and W. Visser. Experiments with Test Case Generation and Runtime Analysis. In E. Börger, A. Gargantini, and E. Riccobene, editors, *Abstract State Machines (ASM'03)*, volume 2589 of *LNCS*, pages 87–107. Springer, March 2003.
4. C. Artho, K. Havelund, and A. Biere. High-Level Data Races. *Software Testing, Verification and Reliability*, 13(4), 2004.
5. C. Artho, K. Havelund, and A. Biere. Using Block-Local Atomicity to Detect Stale-Value Concurrency Errors. In *2nd International Symposium on Automated Technology for Verification and Analysis, Taiwan*, October–November 2004.
6. ASML. `http://research.microsoft.com/fse/asml`.
7. AspectJ. `http://eclipse.org/aspectj`.
8. M. Barnett, K.R.M. Leino, and W. Schulte. The Spec# Programming System: An Overview. In *Construction and Analysis of Safe, Secure, and Interoperable Smart Devices: International Workshop, CASSIS'04*, volume 3362 of *LNCS*, Marseille, France, March 2004. Springer.
9. H. Barringer, A. Goldberg, K. Havelund, and K. Sen. Program Monitoring with LTL in Eagle. In *Parallel and Distributed Systems: Testing and Debugging (PADTAD'04)*, April 2004. Santa Fee, New Mexico, USA.

10. H. Barringer, A. Goldberg, K. Havelund, and K. Sen. Rule-Based Runtime Verification. In *Proceedings of Fifth International VMCAI conference (VMCAI'04)*, volume 2937 of *LNCS*. Springer, January 2004.
11. BCEL. `http://jakarta.apache.org/bcel`.
12. S. Bensalem, M. Bozga, M. Krichen, and S. Tripakis. Testing Conformance of Real-Time Applications by Automatic Generation of Observers. In *Proceedings of the 4th International Workshop on Runtime Verification (RV'04)* [1], pages 19–38. `http://react.cs.uni-sb.de/rv2005`.
13. S. Bensalem and K. Havelund. Deadlock Analysis of Multi-Threaded Java Programs. Internal report, NASA Ames Research Center, April 2003.
14. F. Chen, M. D'Amorim, and G. Roşu. Checking and Correcting Behaviors of Java Programs at Runtime with Java-MOP. In *Proceedings of the 5th International Workshop on Runtime Verification (RV'05)* [1]. `http://react.cs.uni-sb.de/rv2005`.
15. M. Clavel, F. J. Durán, S. Eker, P. Lincoln, N. Martí-Oliet, J. Meseguer, and J. F. Quesada. The Maude System. In Paliath Narendran and Michaël Rusinowitch, editors, *Proceedings of the 10th International Conference on Rewriting Techniques and Applications (RTA-99)*, volume 1631 of *Lecture Notes in Computer Science*, pages 240–243, Trento, Italy, July 1999. Springer-Verlag. System Description.
16. Daikon. `http://pag.csail.mit.edu/daikon`.
17. M. D'Amorim and K. Havelund. Runtime Verification for Java. In *Workshop on Dynamic Program Analysis (WODA'05)*, March 2005.
18. M. D'Amorim and G. Rosu. Efficient Monitoring of Omega-Languages. In *CAV'05*, LNCS. Springer-Verlag, 2005.
19. B. D'Angelo, S. Sankaranarayanan, C. Sanchez, W. Robinson, B. Finkbeiner, H. B. Sipma, S. Mehrotra, and Z. Manna. LOLA: Runtime Monitoring of Synchronous Systems. In *12th International Symposium on Temporal Representation and Reasoning (TIME'05)*, pages 166–174, 2005.
20. D. Drusinsky. The Temporal Rover and the ATG Rover. In *SPIN Model Checking and Software Verification*, volume 1885 of *LNCS*, pages 323–330. Springer, 2000.
21. D. Drusinsky. Monitoring Temporal Rules Combined with Time Series. In *CAV'03*, volume 2725 of *LNCS*, pages 114–118. Springer-Verlag, 2003.
22. D. Drusinsky. Semantics and Runtime Monitoring of TLCharts: Statechart Automata with Temporal Logic Conditioned Transitions. In *Proceedings of the 4th International Workshop on Runtime Verification (RV'04)* [1], pages 2–18. `http://react.cs.uni-sb.de/rv2005`.
23. Eiffel. `http://www.eiffel.com`.
24. Y. Eytani, K. Havelund, S. Stoller, and S. Ur. Toward a Benchmark for Multi-Threaded Testing Tools. *Concurrency and Computation: Practice and Experience*, 2005. To appear.
25. B. Finkbeiner, S. Sankaranarayanan, and H. Sipma. Collecting Statistics over Runtime Executions. In *Proceedings of the 2nd International Workshop on Runtime Verification (RV'02)* [1], pages 36–55. `http://react.cs.uni-sb.de/rv2005`.
26. B. Finkbeiner and H. Sipma. Checking Finite Traces using Alternating Automata. In *Proceedings of the 1st International Workshop on Runtime Verification (RV'01)* [1], pages 44–60. `http://react.cs.uni-sb.de/rv2005`.
27. C. Flanagan and S. Freund. Atomizer: A Dynamic Atomicity Checker for Multithreaded Programs. *SIGPLAN Not.*, 39(1):256–267, 2004.
28. D. Giannakopoulou and K. Havelund. Automata-Based Verification of Temporal Properties on Running Programs. In *Proceedings, International Conference on Au-*

tomated Software Engineering (ASE'01), pages 412–416. ENTCS, 2001. Coronado Island, California.
29. A. Goldberg and K. Havelund. Instrumentation of Java Bytecode for Runtime Analysis. In *Fifth ECOOP Workshop on Formal Techniques for Java-like Programs (FTfJP'03)*, July 2003. Darmstadt, Germany.
30. D. Harel. Statecharts: A Visual Formalism For Complex Systems. *Science of Computer Programming*, 8:231–274, 1987.
31. J. Harrow. Runtime Checking of Multithreaded Applications with Visual Threads. In K. Havelund, J. Penix, and W. Visser, editors, *SPIN Model Checking and Software Verification*, volume 1885 of *Lecture Notes in Computer Science*, pages 331–342. Springer, 2000.
32. K. Havelund. Using Runtime Analysis to Guide Model Checking of Java Programs. In *SPIN Model Checking and Software Verification*, volume 1885 of *LNCS*, pages 245–264. Springer, 2000.
33. K. Havelund and G. Roşu. Monitoring Java Programs with Java PathExplorer. In *Proceedings of the 1st International Workshop on Runtime Verification (RV'01)* [1], pages 97–114. `http://react.cs.uni-sb.de/rv2005`.
34. K. Havelund and G. Roşu. Monitoring Programs using Rewriting. In *Proceedings, International Conference on Automated Software Engineering (ASE'01)*, pages 135–143. Institute of Electrical and Electronics Engineers, 2001. Coronado Island, California.
35. K. Havelund and G. Roşu. An Overview of the Runtime Verification Tool Java PathExplorer. *Formal Methods in System Design*, 24(2), March 2004.
36. K. Havelund and G. Roşu. Efficient Monitoring of Safety Properties. *Software Tools for Technology Transfer*, 6(2):158–173, 2004.
37. K. Havelund and G. Roşu. Synthesizing Monitors for Safety Properties. In *Tools and Algorithms for Construction and Analysis of Systems (TACAS'02)*, volume 2280 of *LNCS*, pages 342–356. Springer, 2002. Best paper award.
38. Jass. `http://csd.informatik.uni-oldenburg.de/~jass`.
39. JML. `http://www.cs.iastate.edu/~leavens/JML`.
40. M. Kim, S. Kannan, I. Lee, and O. Sokolsky. Java-MaC: a Run-time Assurance Tool for Java. In *Proceedings of the 1st International Workshop on Runtime Verification (RV'01)* [1]. `http://react.cs.uni-sb.de/rv2005`.
41. Larch. `http://www.cs.iastate.edu/larch-faq-webboy.html`.
42. Patterns. `http://patterns.projects.cis.ksu.edu`.
43. PSL/Sugar. `http://www.pslsugar.org`.
44. PVS. `http://pvs.csl.sri.com`.
45. RAISE. `http://spd-web.terma.com/Projects/RAISE`.
46. G. Roşu and K. Havelund. A Rewriting-based Approach to Trace Analysis. *Automated Software Engineering*, 12(2):151–197, 2005.
47. S. Savage, M. Burrows, G. Nelson, P. Sobalvarro, and T. Anderson. Eraser: A Dynamic Data Race Detector for Multithreaded Programs. *ACM Transactions on Computer Systems*, 15(4):391–411, November 1997.
48. K. Sen and G. Roşu. Generating Optimal Monitors for Extended Regular Expressions. In *Proceedings of the 3rd International Workshop on Runtime Verification (RV'03)* [1], pages 162–181. `http://react.cs.uni-sb.de/rv2005`.
49. K. Sen, G. Roşu, and G. Agha. Detecting Errors in Multithreaded Programs by Generalized Predictive Analysis of Executions. In M. Steffen and G. Zavattaro, editors, *Formal Methods for Open Object-Based Distributed Systems, 7th IFIP WG 6.1 International Conference, FMOODS 2005*, volume 3535 of *LNCS*, Athens, Greece, June 2005. Springer.

50. Specware. http://www.specware.org.
51. SPIN. http://spinroot.com.
52. T-UPPAAL. http://www.cs.aau.dk/~marius/tuppaal.
53. P. Thati and G. Rosu. Monitoring Algorithms for Metric Temporal Logic Specifications. In *Proceedings of the 4th International Workshop on Runtime Verification (RV'04)* [1], pages 131–147. http://react.cs.uni-sb.de/rv2005.
54. Valgrind. http://valgrind.org.
55. VDM. http://www.csr.ncl.ac.uk/vdm.
56. W. Visser, K. Havelund, G. Brat, and S. Park. Model Checking Programs. In *Proceedings of ASE'00: The 15th IEEE International Conference on Automated Software Engineering*. IEEE CS Press, September 2000.
57. L. Wang and S. Stoller. Run-Time Analysis for Atomicity. In *Proceedings of the 3rd International Workshop on Runtime Verification (RV'03)* [1]. http://react.cs.uni-sb.de/rv2005.