KTH Computer Science
and Communication

Test-inspired runtime verification

Using a unit test-like specification syntax for runtime verification

ADAM RENBERG

Master's Thesis at CSC
Supervisor Valtech: title? Erland Ranvinge
Supervisor CSC: title Narges Khakpour
Examiner: title Johan Håstad

TRITA xxx yyyy-nn

Abstract

Abstract in English. Write when most of the report is written.

Keywords: Runtime Verification, Unit Testing, Program Correctness

Referat

"TODO: Test-inspirerad runtime-verifiering"

Sammanfattning på svenska. Skrivs sist.

Keywords (Sökord? Nyckelord?):

Preface

This is a master thesis / degree project in Computer Science at the Royal Institute of Technology (KTH), Stocholm. The work was done at Valtech Sweden, an IT Consultancy. It was supervised by Erland Ranvinge (Valtech) and Dr. (**todo: check**) Narges Khakpour (CSC KTH).

todo: Thanks to people.. Narges, Erland. Valtech. Proof readers.

Any errors contained in the report are mine and mine alone **(rephrase)** .

Contents

Chapter 1

Introduction

Due to the increasing size and complexity of computer software it has become increasingly difficult, if not impossible, to convince oneself that the software works as desired. This is where verification tools can be used to great effect. Of these tools, testing is the one familiar to most developers, and in wide spread use. The proliferation of agile development practices and test-driven development has also popularized the concept of *unit testing*, in which small modules of a program or system are tested individually.

While testing is popular and often works well, it is incomplete and informal, and thus yields no proof that the program does what it should - i.e. follows its specification. Formal verification techniques, such as theorem proving, model checking (and its bounded variant), can give such proofs. However, they suffer from complexity problems (incompleteness, undecidability) and practical issues, such as the so-called state explosion problem. Often they cannot be fully automated.

A relatively new approach in this area is runtime verification, in which the program *execution* is verified against its specification. With the specification written in a suitably formal language, the program can be given a mathematical proof that its specification is followed.

1.1 Problem Statement

How can runtime verification specifications be written in a manner that uses the syntax of the target program's programming language, and resembles the structure of unit tests?

1.2 Motivation

Checking that a program works correctly is of great interest to software developers. Formal verification techniques are helpful, but as mentioned above, traditional methods can be impractical with larger programs, and verification by testing is in-

formal and incomplete. Runtime verification can here be a lightweight addition to the toolbox of verification techniques.

The specification languages used by runtime verification approaches are often based on formal languages/formalisms (e.g. logic or algebra) and not written in the target program's programming language. This means that writing the specifications requires specific knowledge and expertise in mathematics. It also requires mental context-switching, between writing the program and writing the specification, and special tools to support this specialised language's syntax.

In contrast, unit testing frameworks often utilise the programming language to great effect, and they are a common part of the software development process.

If runtime verification specifications more resembled unit tests, and were written in the target program's programming language, it might popularise the use of runtime verification for checking the correctness of programs.

1.3 Disposition

The rest of this report is structured in as follows. Chapter 2 gives a background to the subject of verifying program correctness. Chapter 3 continues by describing the previous research on runtime verification and the syntax of specification languages. It also gives an overview of the current ideas in unit testing.

Chapter 4 describes the approach this work takes to solving the problem stated in Section 1.1. It describes the syntax, instrumentation and verification techniques used in a proof-of-concept implementation, and gives a formal foundation to a subset of the syntax. Chapter 5 then gives an evaluation of this work... **todo:** .

Conclusions and a discussion of this and future work is done in Chapter 6.

Chapter 2

Background

Runtime verification is a new area of research, but the research on verification and formal approachs goes back several decades. Research of interest include the early work on formal methods, e.g. by Hoare [1] and Floyd [2], and work on logics suitable for runtime verification, e.g. LTL by Pnueli [3]. The seminal work done by Hoare, Floyd and Pnueli lay the foundation for many interesting approaches used for runtime verification. LTL is one of the common formal languages used for specifications in runtime verification.

This chapter gives a short overview of the background to the concepts of this report. It starts with laying out what we mean by proving the correctness of programs in Section 2.1. Section 2.2 describes runtime verification and its place in proving correctness. And finally, Section 2.3 discusses testing - syntax, style and other concepts.

2.1 Proving Correctness

A correctness proof is a certificate, based in mathematics and logics, that a program/system/function follows its specifications, i.e. does what it is supposed to do. There are several approaches, with their respective advantages and disadvantages.

Theorem proving, as started by Hoare [1], Floyd [2] and others, is the manual, semi-automated, or (not so often) fully automated process of mathematically proving that the system follows its specification. There are many ways of doing such proofs.

One way is to prove that at all points in the program, given inputs satisfying some pre-conditions, the outputs will satisfy the post-conditions. By formulating post-conditions for the exit point(s) of the program so that they follow the specification, and by linking together the pre-conditions of program points with their preceding program points' post-conditions, we know that correct indata will yield correct results.

This way of proving correctness often yields the best results. But it is slow, hard to automate, and therefore requires much manual labor. Wading through

large programs thus often becomes impractical.

Model checking is the concept of verifying that a *model* of a system (the *system model*) follows its specification. This requires that both the model and the specification is written in a mathematical formalism. Given this, the task becomes to see if the model satisfies the logical formula of the specification. It is often simpler than theorem proving, and can be automated.

The model of the system is usually structured as a finite state machine (FSM), and verification means visiting all accessible states, checking that they follow the specification (which also can be represented as an FSM). This can be problematic, especially when the state space becomes very big, something known as the *state explosion problem*. There are approaches to address this issue, such as *bounded model checking*, or by using higher-level abstractions.

Proving that a model of a system is correct can be very useful, but it suffers from the inherent flaw of only verifying the model, not the actual system. The model can be difficult to construct, or deviate too far from the system. It can not take the dynamic properties and configuration of the executing code into account.

Runtime verification attempts to solve this by dealing directly with the system, creating its model at runtime.

2.2 Runtime Verification

Runtime verification (RV) is a dynamic approach to checking program correctness, in contrast to the more traditional formal static analysis techniques discussed above.

Runtime verification aspires to be a light-weight formal verification technique, see e.g. [4, 5]. It verifies whether properties of a specification hold *during the execution* of a program.

The specification that should be verified is often written in a formal language, a logic or a calculus, such as linear temporal logic [3]. To build a system model for verifying the properties of the specification, the target program needs to emit or expose certain events and data. The collected events and data are used to build the system model. RV frameworks typically use *code instrumentation* to generate *monitors* for this end.

A monitor is either just part of a recording layer added to the program, which stores the events and data needed for verification, or also the part of the machinery that performs verification.

There are two types of verification: *online* and *offline*. In online verification, the analysis and verification is done during the execution, in a synchronous manner with the observed system. In offline verification, a log of events is analysed at a later time. Online verification allows actions to be taken immediately when violations against the specifications are detected, but with considerable performance cost. Offline verification only impacts the performance by collecting data.

When a violation of the specification occurs, simple actions can be taken (e.g. crash the program, log the error, send emails, etc.), or more complex responses

initiated, resulting in a *self-healing* or *self-adapting* system (see e.g. [6]).

Relevant work on runtime verification include [7], in which Bauer et al. use a three-valued boolean logic (true, false and ?) to reflect that a specification can not only be satisfied (true) or violated (false), but also neither yet, or, in the future it may be either. Bauer et al. also show how they transform specifications into automata (which they call *runtime monitors*).

Bodden presents in [8] a framework for RV implemented through *aspect-oriented programming* using *aspectj*[1] in Java, with specifications written as code annotations. Aspect-oriented programming is described in more detail in Section 3.3.4.

Leucker et al. present a definition of RV in [4], together with an exposition of the advantages and disadvantages, similarities and differences, with other verification approaches. In [5], Delgado et al. classify and review several different approaches and frameworks to runtime verification.

2.3 Testing

On the other end of the program-correctness-checking spectrum is *testing*, which is the practical approach of checking that the program, given a certain input, produces the correct/acceptable output. Testing is not complete (for all but the most trivial programs, it is impossible to write complete tests), and lacks a formal foundation, so it cannot be used for formal verification. Testing can be a complement to more formal techniques, such as RV. It is in many cases the sole correctness-checking tool.

Unit testing is the concept of writing small tests, or test suites, for the units in a program, such as functions, classes, etc. These tests are used during development to test the functionality of the units. The aim is to reduce the risk of breaking existing functionality when developing new features, or modifying existing code, by preventing regression.

Unit testing is quite young, perhaps having begun in earnest in the 90s, and it was popularized by the extreme programming (XP) movement[2]. Testing in general is very old.

Kent Beck introduced the style of the modern unit testing framework in his work on a testing framework for Smalltalk [9]. Together with Eric Gamma he later ported it to Java, resulting in *JUnit*[3]. Today, this has lead to frameworks in several programming languages, and they are collectively called xUnit [10].

Writing unit tests, often using unit testing *frameworks* such as JUnit for Java and *unittest*[4] for Python, is a common practice on many development teams.

Testing is often a manual process, taking up a large part of development time (see e.g. [11]). Still, there are tools to automatically generate tests.

[1] `http://www.eclipse.org/aspectj/`
[2] `http://www.extremeprogramming.org/`
[3] `http://www.junit.org/`
[4] `http://docs.python.org/library/unittest.html`

When discussing testing, and unit testing in particular, we must mention the concept of test-driven development (TDD). Also made popular by XP, it consists of the cycle: (1) write a failing test, (2) make it pass by writing the simplest code you can, and (3) refactor - rewrite the code so that it becomes good. Tests here play the part of specifications for the units of the program.

Chapter 3

Previous Research

As we saw in Section 2.2, runtime verification is a technique for verifying a program's compliance against a specification during runtime. These specifications need to be written somehow, which will be discussed in Section 3.1. Approaches for verification are discussed in Section 3.2. For verification to work, during runtime, the program usually needs to be instrumented in such a way that the verification process can access all pertinent data. This is discussed in Section 3.3

The design of unit test syntax is discussed in Section 3.4. The combination of the two, runtime verification and unit testing, will be the main subject in Chapters 4 and 5.

3.1 Specifications

Specifications come in many forms, from the informal ones like "I want it have cool buttons", to the contractual ones written between companies and their clients, to tests, and to formal specifications, written in formal languages, specifying properties that should verifiably hold for the program. It is these last two types of specifications that we are interested in here, and which play an important role in runtime verification.

In general, specifications should be abstract, written in a high-level language, and succinctly capture the desired property. Writing erroneous specifications is of course a possibility; specifications need to be easier for humans to verify than the program's implementation. There is little point in having a specification as complex as the program itself, except for as a point of reference. A program can be seen as an all-encompassing, perfect, always-true, specification of itself.

3.1.1 Formalisms for Specifications

There are several common formalisms for writing specifications, and many papers that expand, rephrase and illuminate on them. Although they can be quite different, they share a common origin in the work done by Floyd [2], Hoare [1], and

others before them. Floyd thought of formulas specifying in/out properties of statements, and chaining these together to form a formal proof for the program. Hoare elaborated on this idea by basing his proofs on a few axioms of the programming language and target computer architecture, and building the proof from there.

Linear Temporal Logic

Linear Temporal Logic (LTL) was first discussed by Pnueli in [3], and has since been popular in many areas dealing with a system model containing a temporal dimension. As Pnueli describes it, it is simpler than other logics, but expressive enough to describe many problems of interest for verification. This has been affirmed by the diverse use of LTL by many researchers.

LTL uses a system model of *infinite execution traces*, or *histories*, of the states of the execution. LTL specifications are formulas that operate on these states. An LTL formula consists of *propositional variables* that work on the domain model of the state (checking variables, inputs, global state, etc.), the normal logical operators such as negation and disjunction, and some temporal operators. The most basic and common temporal operators are $\boldsymbol{X}$, *next*, and $\boldsymbol{U}$, *until*. Other operators can be derived from these, such as $\boldsymbol{G}$, *globally*, and $\boldsymbol{F}$, *eventually*.

An example LTL formula, taken from a list of common specification patterns [12], could be: S precedes P, i.e. if the state P holds sometime, the state S will hold before it. This is shown in Figure 3.1.

$$\boldsymbol{G}\, P \rightarrow (\neg P\, \boldsymbol{U}\, (S \wedge \neg P)$$

Figure 3.1. An example of an LTL formula. This can be read as: Globally, if P holds, then, before P, S held at some point.

In [7] Bauer et al. introduce a three-valued boolean semantics for LTL, calling it LTL_3, which takes the values (true, false and ?). This logic is arguably more suited for the finite nature of runtime verification, whereas LTL was designed with infinite traces in mind. The semantics of LTL_3 reflect the fact that when verifying runtime verification specifications, the result can not only be that the specification is satisfied or violated; it can be inconclusive as well. For satisfied or violated specifications, no further verification is required - we already know the outcome. For inconclusive results, we need to continue with the verification, as, with future events, the result could change into either satisfied or violated.

There is a counterpart to LTL in the real-time setting called Timed Linear Temporal Logic (TLTL). It introduces clocks to make specifications of real-time properties possible. It is of great interest to runtime verification, but will not be discussed further here. See e.g. [7] for more.

Design by Contract

Design by Contract was introduced by Bertrand Meyer in [13], and has been fully implemented in the Eiffel programming language. A contract is the idea that functions, and methods on objects, promise to fulfill certain post-conditions (or promises) if the inputs they are given fulfill the pre-conditions (or requirements) specified in the contract. Design by Contract also contains constructs for specifying loop-invariants and class-invariants, properties that should always hold during loops and for objects of a class, respectively. Assertions (see below) are also usually available.

Design by Contract is inspired by Hoare logic, and is essentially Hoare logic written in a certain style.

Assertions

A common construct that is part of many popular programming languges, like C, Java and Python, is the `assert` statement. It is a way to state that some predicate should hold at a point in the program. Usually the predicate is an expression in the programming language, and is not supposed to alter the program state.

Assertions are distinct from the normal program flow, and not to be conflated with exceptions. Assertions check for properties that should always be true, anything else would be a programming error.

3.1.2 Writing Specifications

For verification in general, specifications can be written and used externally to the program. They can be used in specialized model-checking tools, in tools for theorem proving, etc.

Runtime verification requires that the specifications are accessible when building and running the program. At the very least, the program needs to be instrumented to expose the correct system model so that the specification can be verified. It is sometimes desired in runtime verification to do online verification, and then the specifications need to be available and embedded into the system. A few different approaches have been tried to support this.

Approaches to writing specifications can be divided into two parts: those that require you to manually mark code for verification, and those that inject the verification code from external specifications.

Rosuenblum [14] uses specially annotated comments, written directly in the code. Bodden [8] uses Java annotations, which are written at function and variable definitions, to mark code for verification. The programming language Eiffel has full language support for Design by Contract, with pre- and post-conditions, invariants, and more. These are written in direct proximity to the code to be verified. For simple cases it is common to write assertions in the program [15].

Other approaches, such as the ones taken by Jalili et al. in [16] and Barringer et al. in [17], use external specification files.

3.2 Verification against Specifications

Specifications for runtime verification are written so that programs can be verified against them - to see whether they follow the specification, or violate parts of it.

There are several ways to verify a program against its specification. A common one, used in [7, 8, 16, 17] among others, is to generate state machines from the specification. These state machines, sometimes called *runtime monitors*, operate with the input language of events emitted by the program.

todo: Write more on this.

3.3 Code Instrumentation

For verification to work, the verifier needs access to events happening in the program. Such events can be functions called, statements executed, variables assigned, etc., depending on the system model of the specification language. The program needs to be instrumented for it to emit such events. This often means wrapping function calls and variable assignments in a "recording layer", which performs the desired action after logging the event. The events can then be passed on to the verification tools.

There are four major approaches used for program instrumentation.

3.3.1 Pre-processing the Code

Rosenblum [14] uses a pre-processor step in the C compilation setup to instrument code, where the specifications (called assertions by Rosenblum) are transformed from comments into regular C code. The verification code is then compiled together with the program.

3.3.2 Post-processing the Code

It is also possible to rewrite the compiled program, instrumenting the code after compilation. This way, the program needs no knowledge of the verification framework. Depending on the compiled objects, this can be more or less difficult. Binary executables and intermediate formats, such as Java Bytecode or Common Intermediate Language for the Common Language Infrastructure used by .Net, require somewhat different approaches.

3.3.3 Dynamic Code Rewriting

In many dynamic languages, such as Python, Ruby or Javascript, it is possible to rewrite the code during runtime, which is sometimes called *monkey patching*. A function to be monitored could be rewritten, adding a lightweight wrapper that records all calls to it, and then passes on the call to the actual function.

3.3.4 Aspects

An interesting approach to code instrumentation is to use aspect-oriented programming. In aspect-oriented theory, a program should be divided into modules, each only dealing with their own *concern*. Logging, however, is a *crosscutting concern*, as it is used by several unrelated modules. The goal is to not scatter logging code across the modules, and to not tangle it with the modules' own logic. This can be done by defining the logging code as *aspects*, which consists of the logging code, called the *advice*, and a *point cut*, which is a formula describing when the advice should be executed. The possible execution points for a point cut are called *join points*. AspectJ[1] is the canonical framework for aspect-oriented programming.

Runtime verification is a typical case of a cross-cutting concern. Bodden [8] uses AspectJ in his runtime verification implementation.

Aspects in AspectJ are implemented as a post-processing step in the compilation process, adding wrapper code for handling the aspects.

3.4 Unit Testing

We discussed testing and unit testing in general in Section 2.3. Here we'll discuss how it works, and what the syntax is like.

3.4.1 xUnit

The xUnit style of unit testing [10] has given rise to unit testing frameworks for many programming languages. Their structure are all based on the same concept, and since JUnit is the canonical implementation, and one of the first, implementation, we will use it for a short demonstration. See Figure 3.2.

In JUnit, and xUnit, you run a *test suite* of *test cases*, which contain tests. The example in Figure 3.2, the test suite is implicitly created by JUnit, although it is possible to create it and control it your self. A *test runner* runs the test suite, reporting progress to the user. When the tests are finished, any errors are displayed.

In the example in Figure 3.2 has two tests, and methods to set up and tear down the tests *fixture*. These functions are usually called *setUp* and *tearDown*, respectively, and are called before and after each test. The fixture is the surrounding set of objects (environment) that the object under test requires to work properly.

Test written in this style are traditional unit tests.

3.4.2 Behaviour-driven Development

There is a style of writing tests called behaviour-driven development [18]. It originated from test-driven development, and is built on the idea that the tests you write should test the behaviour of the program. The simplest example is that you write

[1] `http://www.eclipse.org/aspectj/`

```
// required imports removed for brevity

public class TestSomeClass
    extends TestCase {
  private Environment;

  @Before
  public void setUp() {
    // setup the fixture for each test
    Environment = new Environment();
  }

  @After
  public void tearDown() {
    // clean up the fixture, free memory, etc.
  }

  @Test
  public void testSimpleAddition() {
    // use the language assertion construct
    assert 1+1 == 2
      // use JUnit's assertion functions
      assertEquals(4+7, 11)
  }

  @Test
  public void testThatDoWorkReturnsX() {
    // do setup for this test
    Target t = new Target(...);
    // exercise the object under test
    t.doWork(...);
    // do verification
    assert t.getValues() == x;
  }
}
```

Figure 3.2. An example of unit testing syntax, written as a test case for JUnit.

your unit tests after the behaviour you desire, perhaps naming your tests according to "X should do Y". A more radical example is shown in Figure 3.3.

A test runner for the test in Figure 3.3 would parse each scenario, and for each line find a matching unit of code that does what the line describes. This way of

```
+Scenario 1: Account is in credit+
Given the account is in credit
And the card is valid
And the dispenser contains cash
When the customer requests cash
Then ensure the account is debited
And ensure cash is dispensed
And ensure the card is returned
```

Figure 3.3. An example scenario describing a behaviour, as written in BDD. Scenario taken from [18].

writing tests, or describing behaviours, leads to a outside-in, or top-down, way of writing and thinking about your program.

3.4.3 Mocking and Faking

A common issue when writing unit tests is that, to instantiate some object X, or to call some function Y, the program needs access to some other objects/data/-configuration Z. Z might be something simple, which we can easily create in the test. It might also be a network or database connection, or something doing heavy calculation, or just something complex.

One way to work around this is to create fake/mock/dummy objects. A fake network connection has the same interface as a real network connection, but calling it does not actually transmit anything anywhere, and it might return pre-defined, hard coded data. Fake objects could save what actions are taken upon them, and the test could then verify that these are according to expectations.

3.4.4 Expectations

Instead of writing fake objects, we can create a mock object and pre-record what actions we expect to be taken upon them. This is called writing *expectations* [19]. A simple example of expectations is shown in Figure 3.4.

Figure 3.4 shows a test of a fictional shop. The test tests only one thing, the fill method of the Order object, but it requires a Warehouse object, for access to the inventory. We supply a mock Warehouse, with expectations on which methods should be called on it, with which arguments and what they should return.

An expectation follows a simple pattern:

- A function, with an optional object, which is expected to be called.
- An invocation count of how often the function is expected to be called.
- Expected arguments for the function call. These can be explicit values, or generic types, or rules defining the acceptable values.

```
public class OrderInteractionTester
    extends MockObjectTestCase {
  private static String TALISKER = "Talisker";

  public void testFillingRemovesInventoryIfInStock() {
    // setup - data
    Order order = new Order(TALISKER, 50);
    Mock warehouseMock = new Mock(Warehouse.class);

    // setup - expectations
    warehouseMock.expects(once())
      .method("hasInventory")
      .with(eq(TALISKER),eq(50))
      .will(returnValue(true));
    warehouseMock.expects(once())
      .method("remove")
      .with(eq(TALISKER), eq(50))
      .after("hasInventory");

    // exercise
    order.fill((Warehouse) warehouseMock.proxy());

    // verify
    warehouseMock.verify();
    assertTrue(order.isFilled());
  }
}
```

Figure 3.4. An example of expectations, written using jMock and JUnit. Example taken from [19].

- The return value and modifications to the global state; what should happen when the function is called.
- When the function call should happen, e.g. in what sequence of function calls, in what global state.

There are two common ways of specifying expectations: recording and explicit specification. Figure 3.4 shows an example of how to explicitly specify expectations.

When recording expectations, you create a mock object and call the expected functions, with expected arguments and return values, in the expected order. Then you set the mock into replay mode, and it will replay the recorded expectations, and verify that they occur correctly.

There are several frameworks for working with expectations, such as jMock[2] for Java, Rhino Mocks[3] for .Net and Ludibrio[4] for Python.

[2] http://www.jmock.org/
[3] http://ayende.com/wiki/Rhino+Mocks.ashx
[4] https://github.com/nsigustavo/ludibrio/

Chapter 4

Approach

This chapter describes the proof-of-concept implementation of this report.

4.1 Introduction

As stated in Section 1.1, the objective of this thesis is to investigate whether it is possible to do runtime verification with specifications written in the target program's programming language, structured similar to unit tests. To find a solution for this, there are four issues we need to address:

1. How should the syntax for the specifications be defined, so that it looks similar to that of unit tests, but works for runtime verification? Which language should be used? Which unit testing framework to take inspiration from?

2. How should the program be instrumented to monitor the system, to expose the appropriate events and data, and to build the system model?

3. How will this be used to verify the system against the specification? Online or offline verification? E.g. which techniques should be used to verify the monitored system against the specification?

4. How can the resulting approach be provided with a formal foundation?

This report is a documentation on how to solve these issues. The following sections are each dedicated to one issue, and shows a proof-of-concept of these ideas. The implementation, called *pythonrv*, can be found online[1].

4.1.1 Definitions

Here follows some definition that will be used in the following sections.

[1] `https://github.com/tgwizard/pythonrv`

- A *specification* is an construct that determines the correct behaviour of a program. It could be a document, describing the programs functionality, or a set of inputs and outputs, describing the correct results of the program's computation on that set. It could be a reference implementation[2]. A *formal specification* is a mathematical construct that can be used in verification proofs to show that a program works correctly, i.e. according to its specification.
- *Instrumentation* is the act of rewriting, intercepting, or patching the program to gain access to its internal state and execution flow.
- In *pythonrv* a *specification function* is a Python function describing a specification, which *pythonrv* can use for verification of the program.
- A specification function *monitors* points (functions) of the program, and the points being monitored are called *monitorees*.

4.1.2 Choice of Language

During the development of this proof-of-concept, the biggest factor in deciding what language to use was how it would assist in instrumentation. Instrumentation is discussed in Section 4.3. The language should also be in wide use, support quick development, and have an active testing culture.

Easy access to a non-trivial and actively used system for real-world testing would be a plus. More on this in Chapter 5.

Python[3], among several languages, fits these criteria, and was chosen as the implementation language.

4.2 Syntax

The canonical framework for doing unit testing in Python is the *unittest* framework that is included in all modern versions of Python. Not much development has happened on it in the last years. Many new frameworks have have spawned, such as PyUnit, Nose and py.test. They build upon the style of unittest and mostly add new miscellaneous features, such as better test reporting. The original structure of the unit tests is still prevalent - unittest builds on the xUnit style of unit testing, discussed in Section 3.4.

The next section will illustrate the syntax of *pythonrv*.

4.2.1 Three Examples

The example in Figure 4.1 shows the basics of a *pythonrv* specification, written as a specification function. Line 1 imports the `rv` module from the `pythonrv` package.

[2]For instance, the only specification for python is the canonical CPython implementation. Python is defined as "what CPython does".

[3]`http://www.python.org`

```
1  from pythonrv import rv
2  import fibmodule
3
4  @rv.monitor(func=fibmodule.fib)
5  def spec(event):
6    assert event.fn.func.inputs[0] > 0
```

Figure 4.1. A specification that monitors the function `fib` in the module `fibmodule`. The monitored function is, locally to the specification function, aliased as `func`. The specification asserts that the first input to the monitored function is always greater than zero.

On line 2 it imports the module containing the function to be monitored. Line 5 defines the specification as an ordinary Python function called `spec`, taking one argument, `event`. The instrumentation is done line 4 by using the *function decorator*[4] `rv.monitor`. `rv.monitor` declares that the function `fib` in `fibmodule` should be monitored, and, whenever `fib` is called, `spec` should be called as well.

The specification function itself consists of any valid Python code. It is passed a special argument, `event`, which gives the specification function access to data about the current event. On line 6, the array of input arguments used to call `fib` is accessed to check that the first argument is greater than zero.

The specification function in Figure 4.1 will be called upon every invocation to `fibmodule.fib`.

Figure 4.2 shows how a specification function can monitor two functions. The specification function will be called whenever either of the monitored functions are called. Which function was called can be determined from the `event` argument, as is done on lines 7 and 14. It is the `called` attribute of a function in the `event.fn` structure that allows for this.

The example also shows how the specification can access a history of previous events - events that it has handled in the past. `event.history` is a list of all events that has occurred that this specification monitors. The last element is the current event, and the next-to-last element is the previous element, which can also be accessed as `event.prev`.

Figure 4.3 shows a more advanced example, in which the `next` function of the `event` argument is used. `event.next` allows the specification function to add more specification functions (possibly implemented as closures or lambdas) to be executed when the next event occurs.

On line 9 the function `followup` is added to be executed on the next event. Since `followup` is added in this way - as a "oneshot" specification function - it needs to add itself using `next` for verification on subsequent events. This is done on line 22.

[4]See Section 4.3 for an explanation of function decorators and `rv.monitor`.

```
1   from pythonrv import rv
2   import mymodule
3
4   @rv.monitor(foo=mymodule.foo,
5     bar=mymodule.bar)
6   def spec(event):
7     if event.fn.foo.called:
8       # the foo function was called
9       # either the size of the event history
10      # is 1 - this is the first event - or
11      # the previous event was a call to bar
12      assert len(event.history) == 1 \
13        or event.prev.fn.bar.called
14    elif event.fn.bar.called:
15      # the bar function was called
16      # assert that previous event
17      # was a call to foo
18      assert event.prev.fn.foo.called
```

Figure 4.2. A specification that monitors two functions, `mymodule.foo` and `mymodule.bar`. It asserts that calls to the two functions alternate; that no two calls to `foo` occurs without a call to `bar` in between, and vice versa. The first call has to be to `foo`.

Figure 4.3 also shows how a specification function can turn its verification off - i.e. unsubscribe from future events. `event.finish` and `event.success` are essentially the same, and unsubscribes without further errors. `event.failure` can be thought of as a combination of `event.finish` and `assert False` (which always fails).

4.2.2 Capabilities and Limitations

The examples above show the main capabilities of *pythonrv* specifications. A few minor details were left out, such as how to specify how much history should be saved for a specification, or how to label specifications with error levels, so that different actions can be taken depending on which specification function fails. This is described on the website for *pythonrv*.

4.3 Instrumentation

The previous section showed how *pythonrv* specification functions can be written. This section will describe how these functions can jack themselves into the ordinary

```
1   from pythonrv import rv
2   import mymodule
3
4   @rv.monitor(foo=mymodule.foo,
5     bar=mymodule.bar, baz=mymodule.baz)
6   def spec(event):
7     if event.fn.foo.called:
8       # add function to be called on
9       # next event
10      event.next(followup)
11      event.finish()
12    else:
13      # verification has failed
14      # similar to assert False
15      event.failure()
16
17  def followup(event):
18    if event.fn.bar.called:
19      event.success()
20    elif event.fn.baz.called:
21      assert event.fn.baz.inputs[0] == True
22      # call this function on next event
23      # as well
24      event.next(followup)
25    else:
26      event.failure()
```

Figure 4.3. A more complex example: A specification function that monitors three functions, `foo`, `bar` and `baz`, and makes sure that `foo` is called first, then any number calls to `bar` with the first argument as `True`, and then finally a call to `bar`. After that, any calls are allowed - the specification function will not be used in verification any longer.

control flow of the program and gain access to the function call events and their arguments and associated state.

Instrumentation is done through the `rv.monitor` *function decorator* in *pythonrv*. A Python function decorator is similar to attributes in .Net and annotations in Java. It is essentially a function that takes in a function and returns a function, possibly modifies it, or uses it in some way (decorates it) in the processs. This is used throughout Python to, for instance, turn functions into static or class methods. Figure 4.4 shows an example function decorator definition, and Figure 4.5 shows how to use it.

```
1  # function_decorator.py
2  # the function decorator
3  def decorator(decoratee):
4    # define the closure ("inner function")
5    def wrapper():
6      print "before",
7      # call the decorated function
8      ret = decoratee()
9      print "after",
10     return ret
11   # return the closure
12   return wrapper
```

Figure 4.4. An example of how to define a function decorator.

```
1  # test.py
2  from function_decorator import decorator
3
4  def func_a():
5    print "a",
6
7  func_a()
8  # output: a
9
10 # decorate func_a
11 func_a = decorator(func_a)
12 func_a()
13 # output: before a after
14
15 # decorate func_b - equivalent
16 # to the decoration of func_a
17 @decorator
18 def func_b():
19   print "b",
20
21 func_b()
22 # output: before b after
```

Figure 4.5. An example of how to use the function decorator from Figure 4.4.

`rv.monitor` first takes arguments specifying what functions should be monitored, and then the specification function itself.

In Python, almost all[5] functions belong to a container of some sort - a class, an object, or a module. In Figure 4.5 the functions `func_a` and `func_b` belong to the module `test` (the module's name is the same as that of the file containing the code). These containers are essentially dictionaries (*dicts* in Python parlance) of key-value pairs, where the keys in this case are function names and the values are objects representing the function code. (There are other types of values in these containers as well, which we can ignore).

The instrumentation in *pythonrv* works as follows. First, a wrapper function is defined for each function to be monitored (for each monitoree). This wrapper function's main purpose is to call the specifications attached to the monitored function, and then call the monitored function itself. The wrapper also does some argument copying and such, to prevent side-effects in the specifications from interfering with the monitored function. The container of the monitored function is then extracted, and the reference to the monitored function is overwritten with a reference to the wrapper. See Figure 4.6 for an overview.

The implementation of the instrumentation code in *pythonrv* is more optimized than this - for instance, only ever one wrapper per monitoree is created, independent of the number of specifications that want to monitor it.

4.4 Verification

In *pythonrv*, verification is quite simple. The specification functions are executable, and executing them on the appropriate events, providing access to the current data, verifies that the specification they represent is followed.

Specification functions notify verification violations, that the specifications are not followed, by raising exceptions of the type `AssertionError`. These exceptions are raised when the `assert` statement fails. They can also be raised manually: `raise AssertionError('error message')`.

The verification is performed online, during the program execution. Specifications are verified for all calls to function they monitor unless they explicitly remove themselves by calling one of `event.finish`, `event.success` and `event.failure` (described in Section 4.2).

4.4.1 Dealing with Errors

Whenever a specification violation occurs, and an `AssertionError` is raised, it is passed to an *error handler*. There are two built-in error handlers: One, the default,

[5]This is not true of closures - function defined inside other functions. These functions cannot be directly referenced or modified from outside the defining function. *pythonrv* does not (as of writing) support monitoring of closures.

```
1   # rv.py
2   def monitor(monitorees, specification):
3     for monitoree in monitorees:
4       # define a wrapper for each monitoree
5       def wrapper(*args, **kwargs)
6         event = create_event(...)
7
8         # call specification
9         specification(event)
10
11        # call the actual function - the
12        # monitoree
13        return monitoree(*args, **kwargs)
14
15      # overwrite the monitoree in its container
16      container = get_container(monitoree)
17      setattr(container, monitoree.name, wrapper)
```

Figure 4.6. An overview of the *pythonrv* instrumentation process, written in pseudo-Python. This is just for illustrative purposes and not how *pythonrv* actually does the instrumentation.

that re-raises the exception, and thus crashes the program[6], and a second which just logs the error message, using the standard Python logging module.

It is possible to write custom error handlers for *pythonrv*. See the website for how.

4.4.2 Offline Verification

The current verification approach in *pythonrv* is to perform it online. This obviously affects the performance of the program under test. Offline verification could be used to mitigate this, removing all overhead but for the required recording layer.

To do offline verification in *pythonrv* the events and their associated data would need to be saved (serialized) and replayed outside the context of the running program.

4.5 Formal Foundation

This is the most difficult part, and I haven't really figured it out yet.

A seemingly insurmountable problem quickly arises when attempting to give a formal foundation to the specifications described in Section 4.2. The specifications

[6]Unless some other part of the program, higher up the call stack, suppresses the exception.

are written as ordinary Python functions and, as such, are difficult to formalize. The Python programming language is rather informal - one implementation of it, CPython, serves as the reference implementation. There are no other specifications or formal semantics for Python[7].

One way to go around this is to define a semantics for a subset of Python. Specifications written in this subset will have a formal semantics, and they will have a way to formally prove their correctness.

4.5.1 Definitions and Terms

We begin with some definitions and terms to make the following sections simpler.

- A *formal specification function* is one of four basic functions described in section 4.5.2, or a set of them combined by composition into a composite function, as described in section 4.5.3.
- Formal specification functions have *composition points* where they can be combined with other formal specification functions. A formal specification function can have zero or more composition points. A composition point can be *open* or *closed* - open composition points can be used in composition, while closed have already been used. Composition points can be *required* or *optional*, where optional is the default
- A formal specification function can be either *complete* or *incomplete*. A complete formal specification function is a valid specification. An incomplete formal specification function is not a valid specification, but can, with composition, become complete. Complete formal specification functions have no open required composition points; incomplete formal specification functions have at least one.
- Composition is described by the $\circ$ operator. Let f, g and h be formal specification functions. Let f have one composition point, g two, labeled a and b, respectively, and h none. A valid composition would be: $s = f \circ ((g \circ_a h) \circ_b h)$. s would be a complete formal specification function, as it is composed of formal specification functions, and it has no open composition points.

4.5.2 Python Subset for Formal Specification Functions

The subset of Python that will be provided with a formal semantics consists of four composable specification functions: *null*, *assert*, *next* and *if*. See Figure 4.7.

The E used in the assertion and if statement (lines 6 and 15) denotes any idempotent, immutable Python expression.

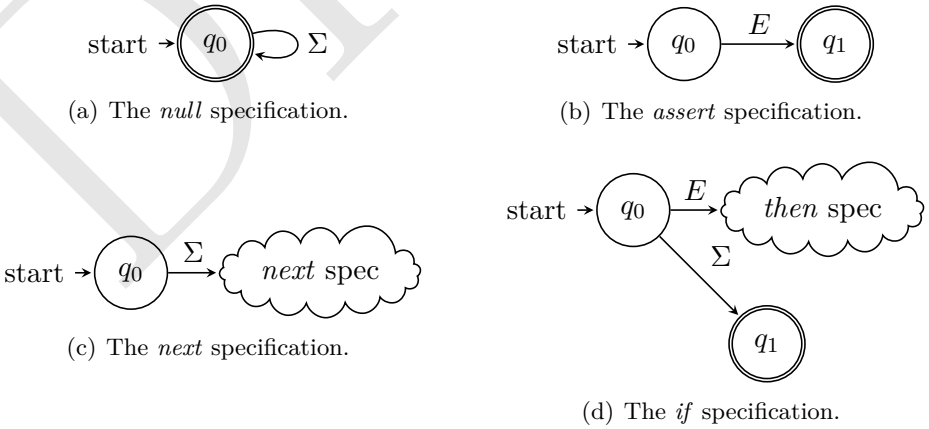
[7]The development of Python is organized mainly through the Python Enhancement Proposal (PEP) process. PEPs are design documents for new features, informally describing their rationales and how they work.

```
1   def null(event):
2     pass
3     # composition point 'after'
4
5   def assert(event):
6     assert E
7     # composition point 'after'
8
9   def next(event):
10    x = # required composition point 'next'
11    event.next(x)
12    # composition point 'after'
13
14  def if(event):
15    if E:
16      # required composition point 'then'
17    else:
18      # composition point 'else'
19    # composition point 'after'
```

Figure 4.7. The four basic formal specification functions.

todo: Give example expressions for *E*?

The four basic formal specification functions correspond to simple finite automata, which are depicted in Figure 4.8. For the system model used, and the semantics behind this, see section 4.5.4. For how to combine the automata, see the next section.



(a) The *null* specification.

(b) The *assert* specification.

(c) The *next* specification.

(d) The *if* specification.

Figure 4.8. The automata for the four basic formal specification functions.

4.5.3 Rules for Composition

There are no implicit precedence rules for the $\circ$ composition operator; parentheses are required.

In this section let *id* be the *null* specification, let a be an *assert* specification, let n be a *next* specifications and let i be an *if* specifications. Let s be any specification. Appending subscripts and $'$ denotes different specifications of the same type.

Composition of the *null* specification

The *null* specification works as an identity function. Composing any function with it, or it with any function, yields the same – see Figure 4.9. Since composition with the *null* specification is the same as applying an identity function, the *after* composition point of the *null* specification inherits the requiredness of the composition point with which it is attached, i.e. attaching the *null* specification to the *then* composition point of the *if* specification will yield an *if* specification with the *then* composition point still required and open.

$$s = s \circ id = id \circ s$$

```
def compose_after(event):
  pass
  s(event)

def compose_before(event):
  s(event)
  pass
```

Figure 4.9. Composition of a *null* specification *id* with an arbitrary specification s.

Due to the identity-function nature of the *null* specification, it will be left out of the composition rules below.

Composition of the *assert* specification

Composing with an *assert* specification a is simple when a is on the left-hand side of the composition. See Figures 4.10, 4.11 and 4.12.

The automata corresponding to the specifications created by the compositions from are quite simple. These are created by just taking the expression E from a and adding it (as part of a conjunctive) to any expressions on all transitions from the start state in the right-hand side automaton.

For when an *assert* specification is on the right-hand side, see the composition rules for *next* and *if* specifications.

$s = a \circ a'$

```
def s(event):
  assert E
  assert E'
```

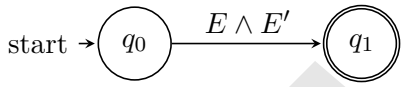


Figure 4.10. A composition of an *assert* specification and another *assert* specification.

$s = a \circ n$

```
def s(event):
  assert E
  next(...)
```

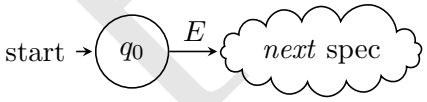


Figure 4.11. A composition of an *assert* specification and an *next* specification, with the *assert* specification on the left-hand side.

$s = a \circ i$

```
def s(event):
  assert E
  if E':
    ...
```

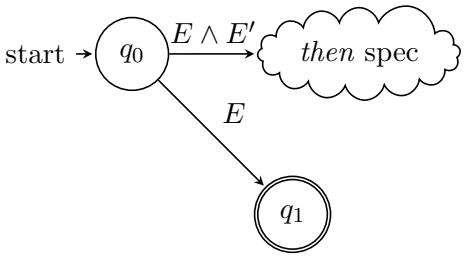


Figure 4.12. A composition of an *assert* specification and an *if* specification, with the *assert* specification on the left-hand side. The optional else clause have been left out, but if it were added, we would draw another transition from $q0$ to the automaton inserted at that composition point labeled $E \wedge \bar{E}'$. $q1$ is the state representing the "fallthrough" of the if statement, at the *after* composition point

Composition of the *next* specification

next specifications have two composition points: one appropriately called *next*, which is required, and one called *after*, which is optional. Compositions using the *after* composition point are reflexive: $n \circ_{after} s = s \circ_{after} n$. For composition rules concerning $a \circ_{after} n$ and $i \circ_{after} n$, see the sections for *assert* and *if* specifications. Also note that:

$$(n \circ_{next} s) \circ_{after} (n' \circ_{next} s') = n \circ_{next} (s \circ_{after} s')$$

As described in section 4.5.2, composition with a *next* specification on the left hand side is simple. It is illustrated again in Figure 4.13.

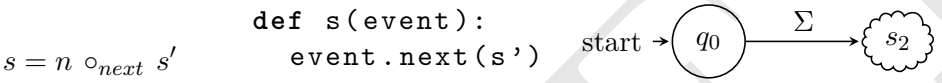


Figure 4.13. A composition of a *next* specification and an *assert* specification, with the *next* specification on the left-hand side.

Composition of the *if* specification

if specifications are the most complex of the four basic specifications, with three composition points: *then*, *else* and *after*, with *then* being required and the other optional. The different combinations with other specifications (*assert*, *next* and *if*) and composition points are quite many, so the figures below only show how each would behave when attached to both the *then* and *else* composition points. The different combinations are then trivial to calculate. See Figures 4.14, 4.15 and 4.16 for the composition rules for *if* specification on the left-hand side. Rules for *if* specifications when they are on the right hand side are defined above.

As can be seen in e.g. Figure 4.14, the "fall-through transition", which leads to the *after* composition point, is labeled as the disjunction of the labels to the *then* and *else* composition points. This is due to the fact that if the expression guarding the *then* composition point evaluates to true, whatever is asserted there must hold. If no formal specification function is attached to the optional *else* composition point, a dummy `assert True` is inserted to generate the proper semantics for the fall-through.

todo: Rethink. if is just adding (conjunction) the E of the if to all transitions in its inner cloud.

4.5.4 Semantics

Qdo.

$s = (i \circ_{then} a_1) \circ_{else} a_2$

```
def s(event):
  if E:
    assert E1
  else:
    assert E2
```

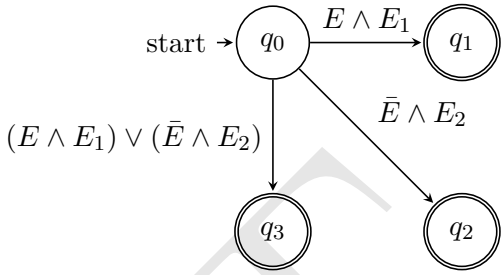


Figure 4.14. A composition of an *if* specification and *assert* specifications, with the *if* specification on the left-hand side.

$s = (i \circ_{then} n_1) \circ_{else} n_2$

```
def s(event):
  if E:
    event.next(n1)
  else:
    event.next(n2)
```

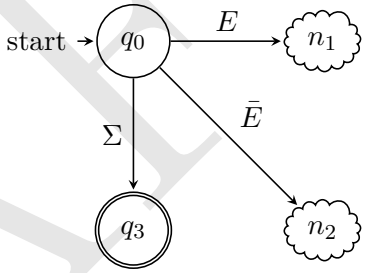


Figure 4.15. A composition of an *if* specification and *next* specifications, with the *if* specification on the left-hand side.

$s = (i \circ_{then} i_1) \circ_{else} i_2$

```
def s(event):
  if E:
    if E1:
      # then 1
    else:
      # else 1
  else:
    if E2:
      # then 2
    else:
      # else 2
```

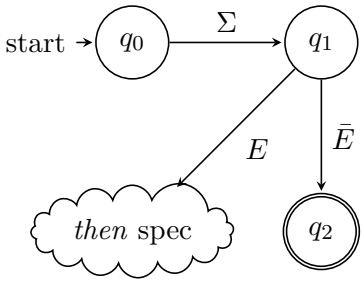


Figure 4.16. A composition of a *next* specification and an *if* specification, with the *next* specification on the left-hand side.

Chapter 5

Evaluation

To see how *pythonrv* would work in a real-world setting it was incorporated into a real-time web application for Valtech Sweden, a medium-sized Swedish company.

The web application is written in Python 2.7 using the Django[1] web framework. It has approximately 10000 lines of code.

There are two questions we need to answer when writing specifications for a program. First, when, in the life-cycle of the program, should we attach the specifications? I.e. when should the code instrumentation be done? And second, and most important: what specifications should be written, and for which functions?

We can answer the first question first. It requires a bit of knowledge on the start up sequence for, and structure of, Django applications.

5.1 Technical Perspective

5.1.1 Anatomy of a Django Application

A Django application follows the Model-View-Controller pattern[2], or as they call it, the Model-Template-View pattern. The model is a representation of the data used by the program, and the templates are the layer that constructs the display for the user. The view links the two together, fetching the correct models for specific requests, and then delegating to the appropriate templates.

Application-specific configuration for Django programs are stored in settings modules, which are ordinary Python files. These contain settings for database connections, authentication, etc.. During startup, Django reads the settings files, starts up its internal machinery, and waits for the first request.

5.1.2 When to Attach

At first glance it might seem desirable to attach the specifications before even starting the Django framework. That way we could monitor the startup process, and all

[1] `https://www.djangoproject.com/`

[2] **todo: Link somewhere?**

of the functionality of Django.

A problem with this, that is due to how Python works, and how *pythonrv* does code instrumentation, is that *pythonrv* needs to load the modules (files) for each function to be monitored. These modules are often heavily dependent on Django, and that it has been started correctly, with all settings loaded.

A suitable time to instrument the program - to enable the specifications - is during startup, after the settings have been loaded. Some specifications, which do not monitor code dependent on the settings, could be loaded before that.

5.1.3 Issues

Early in the process of using *pythonrv* in the program it was discovered that the copying of data, such as function arguments, that *pythonrv* does would not work with Django. The latest version of Django, v1.4.1, uses a module called `cStringIO`, which produces objects that cannot be copied. All functions dealing with web requests are affected by this. This has been fixed in the development branch of Django, but in the meantime, *pythonrv* has an option to disable argument copying, either for all specifications or for a subset of them, to work around this issue.

5.2 Potential Value

Now to the most important question: what specifications could, and should, be written? What value do they provide?

todo: Talk about where specifications would be suitable, and for what.

Chapter 6

Conclusions

This report, and the proof-of-concept implementation *pythonrv*, has shown that it is possible to write specifications in the target programs programming language (Python) and in a manner more similar to unit testing.

However, a few reservations should be mentioned. The specification functions' explicit dealing with time and the actual execution flow leads to some inherent divergences from ordinary unit testing styles.

Also, giving the specifications a formal foundation, and doing formal verification with them, is different, and perhaps more difficult, than with specifications already written in formal languages. The fact that the chosen programming language, Python, does not have a formal semantics defined makes the task quite a bit larger.

The formal foundation given in section 4.5 is thus for a small subset of Python, which makes the math easier, but the resulting semantics less interesting.

If the verification parts of *pythonrv* is unwanted, it could be used as a simple framework for aspect-oriented programming.

6.1 Future Work

The testing tool called expectations, as described in Chapter 3 could fit quite well with the *pythonrv* style of writing specifications.

The performance of the implementation has not been measured or considered in much detail. Benchmark tests for *pythonrv* would be interesting, as would attempts to introduce it as a correctness verification approach for more programs.

Offline verification, discussed in Section 2.2 and Section 4.4 would be interesting.

6.2 Discussion

The trend of software systems in general seems to be toward larger and more complex entities. This makes the automated verification of program correctness, formal or not, ever more important and an essential part of software development. Run-

time verification could have a place there, if it becomes more popular and simpler to integrate and use in ordinary software.

The implementation described in this report, *pythonrv*, is publicly available on the web[1] as free, open source software. People are welcome to try it, incorporate it into their programs, and extend it, as they see fit. With enough interest, *pythonrv* might develop into a mature framework for runtime verification.

[1] `https://github.com/tgwizard/pythonrv`

Bibliography

[1] C. A. R. Hoare, "An axiomatic basis for computer programming," *Communications of the ACM*, vol. 12, pp. 576–580,583, October 1969.

[2] R. W. Floyd, "Assigning meanigs to programs," *Proceedings of Symposium on Applied Mathematics*, vol. 19, pp. 19–32, 1967.

[3] A. Pnueli, "The temporal logic of programs," *Proceedings of the 18th IEEE Symposium on the Foundations of Computer Science (FOCS-77)*, pp. 46–57, 1977.

[4] M. Leucker and C. Schallhart, "A brief account of runtime verification," *The Journal of Logic and Algebraic Programming*, vol. 78, no. 5, pp. 293–303, 2009.

[5] N. Delgado, A. Q. Gates, and S. Roach, "A taxonomy and catalog of runtime software-fault monitoring tools," *IEEE Transactions on Software Engineering*, vol. 30, pp. 859–872, December 2004.

[6] M. C. Huebscher and J. A. McCann, "A survey of autonomic computing degrees, models, and applications," *ACM Computing Surveys*, vol. 40, pp. 7:1–7:28, August 2008.

[7] A. Bauer, M. Leucker, and C. Schallhart, "Monitoring of real-time properties," in *In Proceedings of the 26th Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS), volume 4337 of LNCS*, pp. 260–272, Springer, 2006.

[8] E. Bodden, "Efficient and Expressive Runtime Verification for Java," in *Grand Finals of the ACM Student Research Competition 2005*, March 2005.

[9] K. Beck, "Simple smalltalk testing: With patterns." `http://www.xprogramming.com/testfram.htm`, Retrieved on 2012-07-03.

[10] M. Fowler, "Xunit." `http://www.martinfowler.com/bliki/Xunit.html`, Retrieved on 2012-07-03.

[11] F. P. Brooks, Jr., *The Mythical Man-Month: Essays on Software Engineering.* Addison-Wesley, 1975.

[12] M. B. Dwyer, G. S. Avrunin, and J. C. Corbett, "Patterns in property specifications for finite-state verification," in *Proceedings of the 21st international conference on Software engineering*, ICSE '99, (New York, NY, USA), pp. 411–420, ACM, 1999.

[13] B. Meyer, "Applying "design by contract"," *Computer (IEEE)*, vol. 25, pp. 40–51, October 1992.

[14] D. S. Rosenblum, "A practical approach to programming with assertions," *IEEE Transactions on Software Engineering*, vol. 21, pp. 19–31, January 1995.

[15] D. Bartetzko, C. Fischer, M. Möller, and H. Wehrheim, "Jass – java with assertions," *Electronic Notes in Theoretical Computer Science*, vol. 55, no. 2, pp. 103–117, 2001. RV'2001, Runtime Verification (in connection with CAV '01).

[16] S. Jalili and M. MirzaAghaei, "Rverl: Run-time verification of real-time and reactive programs using event-based real-time logic approach," in *Proceedings of the 5th ACIS International Conference on Software Engineering Research, Management & Applications*, SERA '07, (Washington, DC, USA), pp. 550–557, IEEE Computer Society, 2007.

[17] H. Barringer, A. Goldberg, K. Havelund, and K. Sen, "Rule-based runtime verification," 2003.

[18] D. North, "Introducing bdd," 2006. `http://dannorth.net/introducing-bdd/`, Retrieved on 2012-08-22.

[19] M. Fowler, 2007. `http://martinfowler.com/articles/mocksArentStubs.html`, Retrieved on 2012-08-22.