

Black Box Software Testing

2004 Academic Edition

PART 16 -- ALTERNATIVES TO GUI TEST AUTOMATION

by

Cem Kaner, J.D., Ph.D.

Professor of Software Engineering

Florida Institute of Technology

and

James Bach

Principal, Satisfice Inc.

These notes are partially based on research that was supported by NSF Grant EIA-0113539 ITR/SY+PE: "Improving the Education of Software Testers." Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Kaner & Bach grant permission to make digital or hard copies of this work for personal or classroom use, including use in commercial courses, provided that (a) Copies are not made or distributed outside of classroom use for profit or commercial advantage, (b) Copies bear this notice and full citation on the front page, and if you distribute the work in portions, the notice and citation must appear on the first page of each portion. Abstracting with credit is permitted. The proper citation for this work is "Black Box Software Testing (Course Notes, Academic Version, 2004) www.testingeducation.org", (c) Each page that you use from this work must bear the notice "Copyright (c) Cem Kaner and James Bach, kaner@kaner.com", or if you modify the page, "Modified slide, originally from Cem Kaner and James Bach", and (d) If a substantial portion of a course that you teach is derived from these notes, advertisements of that course should include the statement, "Partially based on materials provided by Cem Kaner and James Bach." To copy otherwise, to republish or post on servers, or to distribute to lists requires prior specific permission and a fee. Request permission to republish from Cem Kaner, kaner@kaner.com.

Black Box Software Testing

Alternatives to GUI-Based Automated Regression Testing

- Several of these slides were developed by Doug Hoffman or in co-authorship with Doug Hoffman for a course that we co-taught on software test automation.
- Many of the ideas in this presentation were presented and refined in Los Altos Workshops on Software Testing.
- LAWST 5 focused on oracles. Participants were Chris Agruss, James Bach, Jack Falk, David Gelperin, Elisabeth Hendrickson, Doug Hoffman, Bob Johnson, Cem Kaner, Brian Lawrence, Noel Nyman, Jeff Payne, Johanna Rothman, Melora Svoboda, Loretta Suzuki, and Ned Young.
- LAWST 1-3 focused on several aspects of automated testing. Participants were Chris Agruss, Tom Arnold, Richard Bender, James Bach, Jim Brooks, Karla Fisher, Chip Groder, Elisabeth Hendrickson, Doug Hoffman, Keith W. Hooper, III, Bob Johnson, Cem Kaner, Brian Lawrence, Tom Lindemuth, Brian Marick, Thanga Meenakshi, Noel Nyman, Jeffery E. Payne, Bret Pettichord, Drew Pritsker, Johanna Rothman, Jane Stepak, Melora Svoboda, Jeremy White, and Rodney Wilson.
- I'm indebted to James Whittaker, James Tierney, Harry Robinson, and Noel Nyman for additional explanations of stochastic testing.

What is automation design?

- Determine the goals of the automation
- Determine the capabilities needed to achieve those goals
- Select automation components
- Set relationships between components
- Identify locations of components and events
- Sequence test events
- Evaluate and report results of test events.

Issues faced in a typical automated test

- What is being tested?
- How is the test set up?
- Where are the inputs coming from?
- What is being checked?
- Where are the expected results?
- How do you know pass or fail?

Automated software test functions

- Automated test case/data generation
- Test case design from requirements or code
- Selection of test cases
- Able to run two or more specified test cases
- Able to run a subset of all the automated test cases
- No intervention needed after launching tests
- Automatically sets-up and/or records relevant test environment
- Runs test cases
- Captures relevant results
- Compares actual with expected results
- Reports analysis of pass/fail

Characteristics of “fully automated” tests

- A set of tests is defined and will be run together.
- No intervention needed after launching tests.
- Automatically sets-up and/or records relevant test environment.
- Obtains input from existing data files, random generation, or another defined source.
- Runs test exercise.
- Captures relevant results.
- Evaluates actual against expected results.
- Reports analysis of pass/fail.

Not all automation is full automation. Partial automation can be very useful.

Capabilities of automation tools

- Automated test tools combine a variety of capabilities. For example, GUI regression tools provide:
 - capture/replay for easy manual creation of tests
 - execution of test scripts
 - recording of test events
 - compare the test results with expected results
 - report test results
- Some GUI tools provide additional capabilities, but no tool does everything well.

Capabilities of automation tools

- Here are examples of automated test tool capabilities:
 - Analyze source code for bugs
 - Design test cases
 - Create test cases (from requirements or code)
 - Generate test data
 - Ease manual creation of test cases
 - Ease creation/management of traceability matrix
 - Manage testware environment
 - Select tests to be run
 - Execute test scripts
 - Record test events
 - Measure software responses to tests (Discovery Functions)
 - Determine expected results of tests (Reference Functions)
 - Evaluate test results (Evaluation Functions)
 - Report and analyze results

Improve testability by providing diagnostic support

- Hardware integrity tests. Example: power supply deterioration can look like irreproducible, buggy behavior.
- Database integrity. Ongoing tests for database corruption, making corruption quickly visible to the tester.
- Code integrity. Quick check (such as checksum) to see whether part of the code was overwritten in memory.
- Memory integrity. Check for wild pointers, other corruption.
- Resource usage reports: Check for memory leaks, stack leaks, etc.
- Event logs. See reports of suspicious behavior. Normally requires collaboration with programmers.
- Wrappers. Layer of indirection surrounding a called function or object. The automator can detect and modify incoming and outgoing messages, forcing or detecting states and data values of interest.

GUI regression is just a special case

- Source of test cases
 - Old
- Size of test pool
 - Small
- Serial dependence among tests
 - Independent
- Evaluation strategy
 - Comparison to saved result

GUI regression is just a special case

- Source of test cases
 - *Old*
 - Intentionally new
 - Random new
- Size of test pool
 - *Small*
 - Large
 - Exhaustive
- Serial dependence among tests
 - *Independent*
 - Sequence is relevant

GUI regression is just a special case

- Evaluation strategy
 - *Comparison to saved result*
 - Comparison to an oracle
 - Comparison to a computational or logical model
 - Comparison to a heuristic prediction. (NOTE: All oracles are heuristic.)
 - Crash
 - Diagnostic
 - State model

A different special case: Exhaustive testing

- MASPAR functions: square root tests
 - 32-bit arithmetic, built-in square root
 - 2^{32} tests (4,294,967,296)
 - 6 minutes to run the tests
 - Much longer to run the oracle
 - Discovered 2 errors that were not associated with any boundary (a bit was mis-set, and in two cases, this affected the final result).
 - 64-bit arithmetic?

A different special case: Exhaustive testing

- MASPAR functions: square root tests
 - Source of test cases
 - Intentional new
 - Size of test pool
 - Exhaustive
 - Evaluation strategy
 - Comparison to an oracle
 - Serial dependence among tests
 - Independent

Random testing: Independent and stochastic approaches

- Random Testing
 - Random (or statistical or stochastic) testing involves generating test cases using a random number generator. Because they are random, the individual test cases are not optimized against any particular risk. The power of the method comes from running large samples of test cases.
- Stochastic Testing
 - Stochastic process involves a series of random events over time
 - Stock market is an example
 - Program typically passes the individual tests: The goal is to see whether it can pass a large series of the individual tests.
- Independent Testing
 - Our interest is in each test individually, the test before and the test after don't matter.

Independent random tests: Function equivalence testing

- Hypothetical case: Arithmetic in Excel
 - Suppose we had a pool of functions that worked well in previous version.
 - For individual functions, generate random number and take function (e.g. log) in Excel 97 and Excel 2000.
 - Spot check results (e.g. 10 cases across the series)
 - Build a model to combine functions into expressions
 - Generate and compare expressions
 - Spot check results
 - For an academic example of this, see the Final Exam for Software Testing 2 (fall, 2003) at <http://blackbox.cs.fit.edu/blog/kaner/archives/000008.html>

Independent random tests: Function equivalence testing

- Hypothetical case: Arithmetic in Excel
 - Source of test cases
 - Random new
 - Size of test pool
 - Large
 - Evaluation strategy
 - Comparison to an oracle
 - Serial dependence among tests
 - Independent

Comparison functions

- Parallel function (Oracle)
 - Previous version
 - Competitor
 - Standard function
 - Custom model
- Computational or logical model
 - Inverse function
 - mathematical inverse
 - operational inverse (e.g. split a merged table)
 - Useful mathematical rules (e.g. $\sin^2(x) + \cos^2(x) = 1$)

Oracles: Challenges

- Incomplete information from oracle
 - May be more than one oracle for SUT
 - Inputs may effect more than one oracle
- Accuracy of information from oracle
 - Close correspondence makes common mode faults likely
 - Independence is necessary:
 - algorithms
 - sub-programs and libraries
 - system platform
 - operating environment

Oracles: Challenges

- Close correspondence reduces maintainability
- Must maintain currency of oracle through changes in the SUT
- Oracle may become as complex as SUT
- More complex oracles make more errors
- Speed of predictions
- Usability of results

Heuristic oracles

- Heuristics are rules of thumb that support but do not mandate a given conclusion. We have partial information that will support a probabilistic evaluation. This won't tell you that the program works correctly but it can tell you that the program is broken. This can be a cheap way to spot errors early in testing.
- Example:
 - History of transactions → Almost all transactions came from New York last year.
 - Today, 90% of transactions are from Wyoming. Why? Probably (but not necessarily) the system is running amok.

Stochastic test: Dumb monkeys

- Dumb Monkey
 - Random sequence of events
 - Continue through crash (Executive Monkey)
 - Continue until crash *or* a diagnostic event occurs. The diagnostic is based on knowledge of the system, not on internals of the code. (Example: button push doesn't push—this is system-level, not application level.)
 - (name coined by Noel Nyman)

Stochastic test: Dumb monkeys

- Dumb Monkey
 - Source of test cases
 - Random new
 - Size of test pool
 - Large
 - Evaluation strategy
 - Crash or Diagnostics
 - Serial dependence among tests
 - Sequence is relevant

Stochastic test using diagnostics

- Telephone Sequential Dependency
 - Symptoms were random, seemingly irreproducible crashes at a beta site
 - All of the individual functions worked
 - We had tested all lines and branches.
 - Testing was done using a simulator, that created long chains of random events. The diagnostics in this case were assert fails that printed out on log files.

Stochastic test using diagnostics

- Telephone Sequential Dependency
 - Source of test cases
 - Random new
 - Size of test pool
 - Large
 - Evaluation strategy
 - Diagnostics
 - Serial dependence among tests
 - Sequence is relevant

Stochastic test: State-model-based

- Testing Based on a State Model
 - For any state, you can list the actions the user can take, and the results of each action (what new state, and what can indicate that we transitioned to the correct new state).
 - Randomly run the tests and check expected against actual transitions.
 - See www.geocities.com/model_based_testing/online_papers.htm

Stochastic test: State-model based

- Testing Based on a State Model
 - Source of test cases
 - Random new
 - Size of test pool
 - Large, medium or small (different substrategies)
 - Evaluation strategy
 - State model or crash
 - Serial dependence among tests
 - Sequence is relevant

Stochastic test: Saved-tests based

- Testing with Sequence of Passed Tests
 - Collect a large set of regression tests, edit them so that they don't reset system state.
 - Randomly run the tests in a long series and check expected against actual results.
 - Will sometimes see failures even though all of the tests are passed individually.

Stochastic test: Saved-tests based

- Testing with Sequence of Passed Tests
 - Source of test cases
 - Old
 - Size of test pool
 - Large
 - Evaluation strategy
 - Saved results or Crash or Diagnostics
 - Serial dependence among tests
 - Sequence is relevant

Another approach to evaluating strategies for automation

What characteristics of the

- goal of testing
- level of testing (e.g. API, unit, system)
- software under test
- environment
- generator
- reference function
- evaluation function
- users
- risks

would support, counter-indicate, or drive you toward a strategy?

- consistency evaluation
- small sample, pre-specified values
- exhaustive sample
- random (aka statistical)
- heuristic analysis of a large set
- embedded, self-verifying data
- state-model-based testing