Dynamic symbolic execution (concolic execution)

Seminar: Understanding of configurable software systems

Bipin Oli

Advisors: Prof. Sven Apel,
Christian Hechtl

Saarland Informatics Campus,
Saarland University

1 Abstract

Concolic execution [7] is a software verification technique that performs symbolic execution together with concrete input values. Concrete values are selected with the help of a constraint solver to guide a program flow in a particular direction. The selection of concrete values helps to scale the verification to a larger program as it makes the symbolic constraints smaller by selecting specific branches in the program. Compared to random execution, this allows us to guide the analysis in a direction likely to have bugs which makes this technique powerful. However, in doing so we sacrifice the completeness of the analysis in favor of the depth of analysis. The sheer number of branches in a large program makes it difficult to perform a complete analysis, so we have to prioritize the branches likely to contribute to finding a bug. There have been a lot of studies to deal with this path explosion problem. In this paper, I have presented state-of-the-art methods to deal with this problem.

2 Introduction

The main idea of concolic testing is to execute the programsimultaneously with concrete values and symbolic values.When the program is executed, symbolic constraints alongthe executed path are collected in a formula calledpath condition. Then, a branch is picked and negated from the pathcondition resulting in a new formula which is then fed to aconstraint solver to check for satisfiability. If it is satisfiable,concrete test inputs are generated to follow the new feasiblepath. If it is unsatisfiable, the new path is infeasible andanother branch has to be picked to be negated. This wayconcolic testing attempts to improve the poor code coverageof random testing. A key characteristic of concolic testing isthat path conditions can be simplified using concrete valueswhenever the decidability of their symbolic constraints goesbeyond the capabilities of the underlying constraint solver.One major problem with concolic testing is that there arein general an exponential number of paths in the programto explore, resulting in the so-calledpath-explosionprob-lem. Recently, several methods have been proposed to at-tack this problem from various angles: using heuristics fo-cused on branch coverage [3], function summaries [8], usingstatic/dynamic program analysis [2] and so on. We pro-pose a new method based oninterpolation, largely complementary to existing approaches, that significantly mitigatespath-explosion by pruning a potentially exponential numberof paths that can be guaranteed to not encounter a bug.

– What is it?
– Where did it start?
– How does it work?
– Give an example
– Why is it important?
– It's contributions
– It's limitations
– Example of the use of this technique in finding bugs in configurable system eg result of SAGE, EXE, etc.

3 Body

3.1 2007: Performing dynamic test generation compositionally

[6, paper]

Gist: The general idea behind this new search algorithm is to perform dynamic test generation compositionally, by adapting (dualizing) known techniques for interproce-dural static analysis to the context of automated dynamic test generation.

Methodlogy: new search algorithm called SMART which stands for duce a new algorithm, dubbed SMARTfor Systematic Modular Automated Random Testing, a more efficent search method then DART without compromising completeness. It tests functions in isolation, collects testing results as function summaries expressed using preconditions on function inputs and postconditionson function outputs, and then re-use those summaries when testinghigher-level functions.

A SMART search performs dynamic test generation composi-tionally, using function summaries as defined previously. Thosesummaries are dynamically computed in a top-down mannerthrough the call-flow graphGPofP. Starting from the top-levelfunction, one executes the program (initially on some random inputs) until one reaches a first functionfwhose execution termi-nates on a return orhaltstatement. One then backtracks insidefas much as possible using DART, computing summaries for thatfunction and each of those DART-triggered executions. When thissearch (backtracking) infis over, one then resumes the originalexecution wherefwas called, this time treatingfessentially asa black-box, i.e., without analyzing it and re-using its previouslycomputed summary instead.

Result: SMART can perform dynamic test generation com-positionally without any reduction in program path coverage. Wealso show that, given a bound on the maximum number of feasiblepaths in individual program functions, the number of program exe-cutions explored by SMART is linear in that bound, while the num-ber of program executions explored by DART can be exponentialin that bound. SMART = scalable DART

3.2 2006: Software Partitioning for Effective Automated Unit Testing

[5]

Gist: present an approach thatidentifies control and data inter-dependencies between soft-ware components using static program analysis, and dividesthe source code into units where highly-intertwined compo-nents are grouped together. Those units can then be testedin isolation using automated test generation techniques andtools, such as dynamic software model checkers

Methodlogy: group together functions or components that share interfaces of complexity higher than a particular threshold. Complexity is determined by the popularity and sharing of code. The idea is that if the function is very popular and is being called from a lot of places then it is likely that it is not closely liked to any component. And, the sharing of code meaning if two functions share many of the same functions then it is likely that the higher level operation they perform is close to each other.

Configurability part: evaluated the effectiveness by applying the algorithm to the open source implementation of oSIP protocol (http://www.gnu.org/software/osip/osip.html) which is a telephony protocol for call establishment.

Result: showing that auto-matic software partitioning can significantly increase testcoverage without generating too many false alarms causedby unrealistic inputs being injected at interfaces betweenunits

3.3 2007: Hybrid concolic testing (**)

Gist: an algorithm that in-terleaves random testing with concolic execution to obtainboth a deep and a wide exploration of program state space.Our algorithm generates test inputs automatically by inter-leaving random testing until saturation with bounded ex-haustive symbolic exploration of program points. It thuscombines the ability of random search to reachdeeppro-gram states quickly together with the ability of concolic test-ing to explore states in a neighborhood exhaustively.

Methodlogy: presenthybrid concolic testing, a simple algorithmthat interleaves the application of random tests with con-colic testing to achieve deep and wide exploration of theprogram state space. From the initial program state, hy-brid concolic testing starts by performing random testingto improve coverage. When random testingsaturates, thatis, does not produce any new coverage points af-ter run-ning some predetermined number of steps, the algorithmautomatically switches to concolic executionfrom the cur-rent program stateto perform an ex-haustive bounded depthsearch for an uncovered coverage point. As soon as one isfound, the algorithm reverts back to concrete mode. Theinterleaving of random testing and concolic execution thususes both the capacity of random testing to inexpensivelygenerate deep program states through long program execu-tions and the capability of concolic testing to exhaustivelyand symbolically search for new paths with a limited looka-head.

The interleaving of random and symbolic techniques isthe crucial insight that distinguishes hybrid concolic testingfrom a naive approach that simply runs random and con-colic tests in parallel on a program. This is because manyprograms show behaviors where the program must reach aparticular statesand then follow a precise sequence of in-put events 'alpha' order to get to a required coverage point.It is often easy to reachsusing random testing, but notthen to generate the precise sequence of events 'alpha'. On theother hand, while it is usually easy for concolic testing togenerate 'sigma', concolic testing gets stuck in exploring a hugenumber of program paths before even reaching the states.

In the end, hybrid concolic testing has the same limita-tions of symbolic execution based test generation: the dis-covery of uncovered points depends on the scalability andexpressiveness of the constraint solver, and the exhaustivesearch for uncovered points is limited by the number ofpaths to be explored. Therefore, in general, hybrid concolictesting may not achieve 100 percent coverage, although it can im-prove random testing considerably. Further, the algorithmis-not a panacea for all software quality issues. While we pro-vide an automatic mechanism for test input generation, allthe other effort required in testing, for example, test oraclegeneration, assertion based verification, and mock environment creation still have to be performed as with any othertest input generation algorithm. Further, we look for codecoverage, which may or may not be an indicator of codereliability.

Configurability part: compare random, concolic, andhybrid concolic testing on the VIM text editor (150K linesof C code) and on an implementation of the red-black treedata structure. Our experiments indicate that for a fixed test-ing budget, hybrid concolic testing technique outperformsboth random and concolic in terms of branch coverage. of the state space exhaustively. In contrast, hybrid concolictesting switches to inexpensive random test-ing as soon as it identifies-someuncovered point, relying onfast random testing to explore as much of the state space aspossible. In this way, it avoids expensive constraint solv-ing to perform exhaustive search in some part of the statespace. Moreover, if random testing does not hit a new cov-erage point, it can take advantage of the locally exhaustivesearch provided by concolic testing to continue from a newcoverage point

3.4 2008: Heuristics for Scalable Dynamic Test Generation

Gist: several such heuristic search strategies, including anovel strategy guided by the control flow graph of the programunder test.

Methodlogy: We propose a search strategy that is guided by the staticstructure of the program under test, namely the control flowgraph (CFG). In this strategy, we choose branches to negatefor the purpose of test generation based on their distance inthe CFG to currently uncovered branches. We experimentallyshow that this greedy approach to maximizing the branchcoverage helps to improve such coverage faster, and to achievegreater final coverage, than the default depth-first searchstrategy of concolic testing.We further propose two random search strategies. While intraditional random testing a program is run on random inputs,these two strategies test a program along random executionpaths. The second attempts to sample uniformly from the spaceof possible program paths, while the third is a variant we havefound to be more effective in practice
have implemented these search strategies in CREST, anopen-source prototype test generation tool for C

Configurability part: We have implemented these strategies in CREST, ouropen source concolic testing tool for C, and evaluated them on twowidely-used software tools, grep 2.2 (15K lines of code) and Vim5.7 (150K lines). On these benchmarks, the presented heuristicsachieve significantly greater branch coverage on the same testingbudget than concolic testing with a traditional depth-first searchstrategy.

3.5 2009: Fitness-Guided Path Exploration in Dynamic Symbolic Execution

Gist: To address the space-explosion issue in path exploration, we propose a novelapproach called Fitnex, a search strategy that uses state-dependent fitness values (computed through a fitness func-tion) to guide path exploration. The fitness function mea-sures how close an already discovered feasible path is toa particular test target (e.g., covering a not-yet-coveredbranch). Our new fitness-guided search strategy is inte-grated with other strategies that are effective for explorationproblems where the fitness heuristic fails.

Methdlogy: The core of our approach is the Fitnex search strat-egy guided by fitness values computed with a fitness func-tion (Section 4.1). To deal with program branches notamenable to a fitness function, our approach includes integration of the Fitnex strategy with other search strategies(Section 4.2)

A fitness function (Section 4.1.1) gives a measurementon how close an explored path is to achieving a test tar-get (e.g., covering a not-yet-covered branch). We computea fitness value for each already explored path and priori-tize these known paths based on their fitness values (Sec-tion 4.1.2). We compute a fitness gain for each branch in theprogram under test and prioritize branching nodes based ontheir corresponding branches' fitness gains (Section 4.1.3).During path exploration, we give higher priority to flippinga branching node with a better (higher) fitness gain in a pathwith a better (lower) fitness value (Section 4.1.4).

3.6 2013: Boosting Concolic Testing via Interpolation

Gist: propose a new and complementarymethod based oninterpolation, that greatly mitigates path-explosion by subsuming paths that can be guaranteed to nothit a bug.

Methodlogy: first, assume that the program is annotatedwith certain bug conditions of the form "if C then bug",where if the conditionCevaluates to true along a path,the path is buggy. Then, whenever an unsatisfiable pathcondition is fed to the solver, an interpolant is generatedat each program point along the path. The interpolant at agiven program point can be seen as a formula thatsuccinctlycaptures the reason of infeasibility of paths at the programpoint. In other words it succinctly captures the reason whypaths through the program point are not buggy. As a re-sult, if the program point is encountered again through adifferent path such that the interpolant is implied, the newpath can besubsumed, because it can be guaranteed to notbe buggy. The exponential savings are due to the fact that not only is the new path subsumed, but also the paths thatthis new path would spawn by negating its branches.

Unfortunately, methods such as [12, 14, 11] cannot be useddirectly for concolic testing due to several challenges. First,the soundness of these methods relies on the assumptionthat an interpolant at a node has been computed after ex-ploring the entire "tree" of paths that arise from the node.In concolic testing, this assumption is invalid as the testercan impose an arbitrary search order. For example, concolictesters such as Crest [3] and KLEE [4] use often many heuris-tics that may follow a random walk through the search space,thus making this method unsound. To address this problem,we need to keep track of nodes whose trees have been ex-plored fully (in which case we say the node is annotated withafull-interpolant) or partially (similarly, ahalf-interpolant).Under this new setting, only nodes with full-interpolants arecapable of subsumption in a sound manner. As a result, theamount of subsumption depends on how often nodes get an-notated with full-interpolants from the paths explored by thecon-colic tester. Unfortunately our benchmarks in Section 6showed that the above method by itself results in very fewnodes with full-interpolants, thereby providing poor bene-fit to the concolic tester, because the tester rarely exploresthe entire tree of paths arising from a node. Hence, an im-portant challenge now is to "accelerate" the formation offull-interpolants in order to increase subsumption.

For this,we introduce a novel technique calledgreedy confirmationthat performs limited path exploration (i.e., execution of afew extra paths) by itself, guided by subsumption, with anaim to produce a full-interpolant at nodes currently anno-tated with a half-interpolant. It is worth mentioning thatthis execution of few paths is done without interfering withthe search order of the concolic tester. This technique ul-timately resulted in a significant increase in subsumptionfor our benchmarks, and is vital for the effectiveness of ourmethod.We implemented our method and compared it with a pub-licly available concolic tester, Crest [3]. We found that forthe price of a reasonable overhead to compute interpolants,a large percentage of paths executed by those heuristics canbe subsumed thereby increasing their coverage substantially.

Result: We attacked the path-explosion problem of concolic test-ing by pruning redundant paths using interpolation. Thechallenge for interpolation in concolic testing is the lack ofcontrol of search order. To solve this, we presented the concept of half and full interpolants that makes the use of in-terpolants sound, and greedy confirmation that acceleratesthe formation of full-interpolants thereby increasing the like-lihood of subsuming paths.

3.7 2014: A Context-Guided Search Strategy inConcolic Testing

Gist: While moststrategies focus on coverage information in the branch selection process, we introduce CGS which considers contextinformation, that is, how the execution reaches the branch.Our evaluation results show that CGS outperforms otherstrategies.

Methodlogy:

Configurability part:

Result:

3.8 ¡¡paper¿¿

Gist:

Methodlogy:

Configurability part:

Result:

4 Different catagorical bodies

see ralated work section of https://dl.acm.org/doi/pdf/10.1145/2635868.2635872 i.e A Context-Guided Search Strategy inConcolic Testing paper

- summerize, give an overview of the main points of each source and combine them into a coherent whole
- Analyze and interpret: don't just paraphrase other researchers—add your own interpretations where possible, discussing the significance of findings in relation to the literature as a whole
- Critically evaluate: mention the strengths and weaknesses of your sources
- Write in well-structured paragraphs: use transition words and topic sentences to draw connections, comparisons and contrasts

5 Conclusion

6 Papers

- 2006: they worked on backtracking algorithms for search heuristics [17]
- 2007: they combined the fuzzing techniques to improve the coverage [11] [6] [8]
- 2008: Heuristic based approach to select the branches [1]
- 2009: they worked on the fitness guided approach to improve the coverage [16]
- 2013: they boosted concolic testing by subsuming paths that are guaranteed to not hit a bug with their interpolation technique [9]
- 2014: they introduced a concept of context guided search strategy [13]
- 2018: automatic selection of suitable heuristic [2]
- 2018: template guided approach [3]
- 2018: based on probability of program paths and the cost of constraint solving [14]
- 2018: they improved the speed of SMT solver by removing the IR layer making it more practical to keep bigger constraints [18]
- 2019: fuzzy search strategy [12]
- 2019: adaptably changing search heuristics [4]
- 2021: Pathcrawler: proposed different strategies to improve the performance of concolic execution on exhaustive branch coverage [15]
- 2022: Dr. Pathfinder [10] combined concolic execution with deep reinforement learning to prioritize deep paths over shallow ones for hybrid fuzzing

References

1. Burnim, J., Sen, K.: Heuristics for scalable dynamic test generation. In: 2008 23rd IEEE/ACM International Conference on Automated Software Engineering. pp. 443–446 (2008)

2. Cha, S., Hong, S., Lee, J., Oh, H.: Automatically generating search heuristics for concolic testing. In: Proceedings of the 40th International Conference on Software Engineering. pp. 1244–1254 (2018)
3. Cha, S., Lee, S., Oh, H.: Template-guided concolic testing via online learning. In: Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering. pp. 408–418 (2018)
4. Cha, S., Oh, H.: Concolic testing with adaptively changing search heuristics. In: Proceedings of the 2019 27th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering. pp. 235–245 (2019)
5. Chakrabarti, A., Godefroid, P.: Software partitioning for effective automated unit testing. In: Proceedings of the 6th ACM & IEEE International conference on Embedded software. pp. 262–271 (2006)
6. Godefroid, P.: Compositional dynamic test generation. In: Proceedings of the 34th annual ACM SIGPLAN-SIGACT symposium on Principles of programming languages. pp. 47–54 (2007)
7. Godefroid, P., Klarlund, N., Sen, K.: Dart: Directed automated random testing. In: Proceedings of the 2005 ACM SIGPLAN conference on Programming language design and implementation. pp. 213–223 (2005)
8. Godefroid, P., Levin, M.Y., Molnar, D.: Sage: Whitebox fuzzing for security testing: Sage has had a remarkable impact at microsoft. Queue 10(1), 20–27 (2012)
9. Jaffar, J., Murali, V., Navas, J.A.: Boosting concolic testing via interpolation. In: Proceedings of the 2013 9th Joint Meeting on Foundations of Software Engineering. pp. 48–58 (2013)
10. Jeon, S., Moon, J.: Dr. pathfinder: hybrid fuzzing with deep reinforcement concolic execution toward deeper path-first search. Neural Computing and Applications 34(13), 10731–10750 (2022)
11. Majumdar, R., Sen, K.: Hybrid concolic testing. In: 29th International Conference on Software Engineering (ICSE'07). pp. 416–426. IEEE (2007)
12. Sabbaghi, A., Rashidy Kanan, H., Keyvanpour, M.R.: Fsct: A new fuzzy search strategy in concolic testing. Information and Software Technology 107, 137–158 (2019), https://www.sciencedirect.com/science/article/pii/S0950584918302428
13. Seo, H., Kim, S.: How we get there: A context-guided search strategy in concolic testing. In: Proceedings of the 22nd ACM SIGSOFT International Symposium on Foundations of Software Engineering. pp. 413–424 (2014)
14. Wang, X., Sun, J., Chen, Z., Zhang, P., Wang, J., Lin, Y.: Towards optimal concolic testing. In: Proceedings of the 40th International Conference on Software Engineering. pp. 291–302 (2018)
15. Williams, N.: Towards exhaustive branch coverage with pathcrawler. In: 2021 IEEE/ACM International Conference on Automation of Software Test (AST). pp. 117–120 (2021)
16. Xie, T., Tillmann, N., De Halleux, J., Schulte, W.: Fitness-guided path exploration in dynamic symbolic execution. In: 2009 IEEE/IFIP International Conference on Dependable Systems & Networks. pp. 359–368. IEEE (2009)
17. Yan, J., Zhang, J.: Backtracking algorithms and search heuristics to generate test suites for combinatorial testing. In: 30th Annual International Computer Software and Applications Conference (COMPSAC'06). vol. 1, pp. 385–394. IEEE (2006)
18. Yun, I., Lee, S., Xu, M., Jang, Y., Kim, T.: QSYM: A Practical Concolic Execution Engine Tailored for Hybrid Fuzzing. In: Proceedings of the 27th USENIX Security Symposium (Security). Baltimore, MD (Aug 2018)