

Hybrid Reachability Analysis Using Static Graphs And Dynamic Execution Traces For Oss Vulnerabilities

Purv Rakeshkumar Chauhan
Arizona State University, Tempe, AZ
purvchauhan@gmail.com

ABSTRACT

The extensive adoption of third-party libraries, along with the rapid detection of new vulnerabilities, compounds the security issues in the open-source software ecosystem. Although traditional static and dynamic analyses provide critical information about vulnerability reachability, they often end up having limited applicability due to false positives or lack of actual area coverage. This paper presents a new metric known as hybrid reachability analysis, which can be leveraged to improve the precision of risk impact assessments. The new approach integrates dynamic execution traces with static program graphs. The practical utility of this hybrid approach for improving accuracy, reducing false positives, and supporting reasonable priorities for vulnerabilities is also shown through experimentation. Furthermore, we present a tri-fold synthesis of the literature around existing studies of relevance to reachability analyses and hybrid approaches. We conclude by arguing that combining graph and trace data can foster future vulnerability management workflows, and a conceptual model is offered to show how hybrid analysis can advance more sustainable software development. Other limitations covered are scalability and runtime overhead, as well as promising future work such as distributed tracing, embedded machine learning, and extending methods to support continuous deployment. This article provides a comprehensive analysis of the challenges associated with OSS undertaken from traditional analyses and hybrid methodologies.

Keywords: Hybrid Reachability Analysis, Open-Source Software Security, Static Graphs, Dynamic Execution Traces, Vulnerability Detection, Dependency Risk Analysis, Software Assurance.

1. INTRODUCTION

Open-source software (OSS) plays a critical role in software development today, as it has the potential to reduce costs, protect against vendor lock-in, support rapid innovation, and ultimately lead to quicker time to market. In fact, the code and running software we use almost always have a high level of incorporation of shared code, and empirical research reflects this evidence with studies that have shown OSS present in everything from enterprise applications to critical infrastructures (Gkortzis et al., 2021). Yet, this applied omnipresence creates the challenge of attacks, as the security holes discovered in OSS are

rapidly propagated into software supply chains, which can have costly security consequences (Zimmermann, 2019). In addition, deeply nested dependencies mean that there is a lot of code, which is easy to exploit, existing multiple levels deep in the main code base (Pashchenko et al., 2018). The goal of reachability analysis - whether a vulnerable function or code segment can be executed in an application context - is thus an important notion in evaluating exploitable vulnerabilities. Static approaches generate program representations that explore multiple potential operational paths such as conferring call graphs, control-flow graphs, and dependency graphs (Smaragdakis et al., 2011). While static techniques assist in early identification of vulnerabilities, they often leads to overestimation and false-positives. Dynamic techniques, on the other hand, monitor logs and execution traces at runtime, offering high accuracy by verifying which paths are (and are not) taken in practice; however, test coverage, and thus, diversity of input limits their effectiveness, when paths, especially critical paths, are not exercised, leading to false negatives (Bouajjani and Touili, 2012).

The hybrid paradigm to increase the accuracy of a vulnerability detection substitutes a graph-based model for execution traces to have static breadth and dynamic precision. Recent research has demonstrated that hybrid approaches can alleviate the incompleteness of a dynamic-only approach and exhibit a considerable decrease of false positives when compared to static-only approaches (Plate, 2015). The isolation of a pure method is not sufficient because of the sheer size and diversity of dependencies in OSS ecosystems. Additionally, hybrid reachability frameworks are increasingly seen as enabling technologies for fields outside of software engineering, like software reliability in renewable energy platforms, AI-driven security systems, and cybersecurity for critical infrastructures—all of which significantly rely on OSS components (Shahzad et al., 2012).

Despite the progress, still a great deal of issues remain. Firstly, it is hard to get precision without losing scalability at the same time, especially in large open-source software projects with thousands of libraries that are interdependent (Goseva-Popstojanova & Perhinschi, 2015). The other problem related to hybrid methods is that dynamic trace coverage is incomplete, particularly for paths

that are rarely executed or are specific to a certain environment. The issues that remain include the combination of coarse-grained vulnerability disclosures (like CVEs) with fine-grained reachability at the function level, and management of dynamic language features such as coverage and runtime code loading. On top of all that, the lack of standardized benchmarks and evaluation metrics makes it difficult to do a thorough comparison between tools and techniques.

One of the main aims of this paper is to carry out a systematic evaluation of hybrid reachability analysis concerning OSS vulnerabilities. It first discusses the theoretical basis of static, dynamic, and hybrid methods before surveying representative research contributions and toolchains, and drawing attention to their design trade-offs, empirical performance, and limitations. The paper presents new opportunities and possible ways forward but also highlights research issues and gaps that remain open. This paper aims to combine knowledge from different disciplines to expose the current situation and lead to increased methods and analysis for future, more precise, scalable, and efficient methods for discovering the reachability of OS vulnerabilities via vulnerability analysis of OSS.

2. LITERATURE REVIEW

A multitude of scholars have studied software vulnerability in static, dynamic, and hybrid forms of vulnerability testing, which each have a differing perspective on impact assessment and reachability. Graph-based static methods have become popular due to their scalability and thorough coverage; however, false positives are frequent due to over-approximation (Livshits & Lam, 2005; Newsome & Song, 2005). Through the use of runtime execution traces, dynamic analysis has enhanced accuracy; however, its success is mainly determined by the design of the workload and the completeness of inputs (Cadard et al., 2008; Godefroid et al., 2005). In order to improve vulnerability detection and decrease noise, hybrid approaches have recently been proposed that combine the precision of dynamic execution data with the breadth of static analysis (Arzt et al., 2014; Enck et al., 2014). Additional studies have extended these foundations by focusing on dependency management, exploitability assessment, and integration of hybrid techniques into practical security workflows that are summarized as key contributions in the following table.

Table 2.1. Summary of Key Approaches in Static, Dynamic, and Hybrid Vulnerability Analysis

Focus	Findings (Key results and conclusions)	Reference
Static source-level security analysis for Java web applications.	Presented a static analysis framework that detects common web vulnerabilities (SQL injection, XSS, HTTP splitting) in Java codebases with an IDE auditing UI. Demonstrated practical value by finding multiple real vulnerabilities in popular open-source Java apps and highlighted tradeoffs between precision and developer usability.	(Livshits & Lam, 2005)
Dynamic taint analysis for automatic exploit detection (binary-level).	Introduced dynamic taint tracking (TaintCheck) to detect overwrite/exploit attacks at runtime without source code. Demonstrated high detection accuracy (with almost no false positives in the program studied) and automatic signature generation capability, although it noted overhead runtime cost and tradeoffs in deployment.	(Newsome & Song, 2005)
Directed automated random testing (DART)—combining static interface extraction with dynamic testing.	Proposed an automated framework that statically extracts program input interfaces then drives dynamic test executions guided by solving constraints. Demonstrated an effective automated generation of tests that improves coverage and detects security bugs, demonstrating the benefit of systematically exploring paths by combining static and dynamic techniques.	(Godefroid et al., 2005)
Symbolic execution engine for high-coverage test generation (KLEE).	Showed an automated symbolic execution tool that can produce high-coverage tests for complex system programs (e.g., GNU coreutils). Reported significant code coverage and bug-finding capabilities, and the problem of scalability and environment modeling, which would be a basis for hybrid (static+dynamic) solutions.	(Cadard et al., 2008)
System-wide dynamic taint tracking for Android (TaintDroid).	Produced a practical runtime information-flow tracking system for Android that tracks many taint sources with low overhead in real time. Evaluations uncovered numerous privacy leaks in third-party apps, and elucidated strengths and weaknesses of dynamic tracing in real-world mobile ecosystems.	(Enck et al., 2014)

Precise static taint analysis for Android apps (FlowDroid).	In this paper we present a context, flow, field and object sensitive static taint analysis targeted at Android applications. We show improved precision compared to existing static approaches and reduced false positives for privacy/leak detection; we also state the limitations of static in this dynamic context and the advantages of using runtime evidence.	(Arzt et al., 2014)
Hybrid vulnerability discovery: augmenting fuzzing with selective symbolic execution (Driller).	Showed that combining inexpensive fuzzing with selective concolic execution finds deeper memory-corruption bugs than fuzzing or symbolic execution alone. Demonstrated practical gains on real binaries and argued for selective hybrid orchestration to mitigate path explosion while improving reachability.	(Stephens et al., 2016)
Practical concolic execution engine optimized for hybrid fuzzing (QSYM).	Presented an efficient concolic engine designed to integrate with fuzzers; achieved substantial performance improvements enabling hybrid fuzzing at scale. Uncover new bugs in actual software implementations and discuss engineering trade-offs that move hybrid reachability strategies.	(Yun et al., 2018)
Efficient mutation-based fuzzer with techniques to solve path constraints (Angora).	Introduced mutation strategies and scalable byte-level taint tracking to more effectively solve branch constraints without heavy symbolic execution. Demonstrated significant improvements in bug discovery and branch coverage, illustrating alternative hybrid-like solutions that reduce dependence on heavyweight static/symbolic components.	(Chen & Chen, 2018)
Hybrid static + dynamic analysis platform for Android vulnerability detection (AndroShield).	An applied hybrid toolchain was discussed that combined static analysis (sources/sinks, FlowDroid integration) with dynamic monitoring and a web reporting frontend. The assessments not only uncovered drawbacks such as information leakage and insecure network request but also highlighted the practical integration and limitations linked with dynamic coverage and mapping alerts to specific code paths.	(Amin et al., 2019)

3. PROPOSED THEORETICAL MODEL FOR HYBRID REACHABILITY ANALYSIS

Conceptual Overview

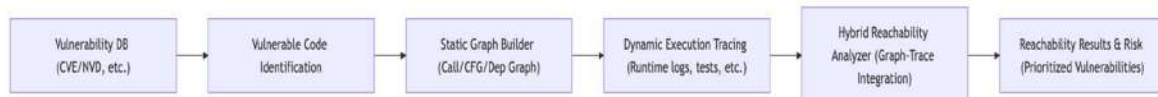
The hybrid reachability analysis framework, in contrast, merges static graphs and dynamic execution traces to analyze whether the vulnerabilities of open-source software (OSS) dependencies that are already known can be exploited by attackers in reality. The model is constructed using two inputs that work in concert:

- **Static graph building** from the application's and the dependencies' codebases, producing call graphs, control-flow graphs, and dependency graphs. This

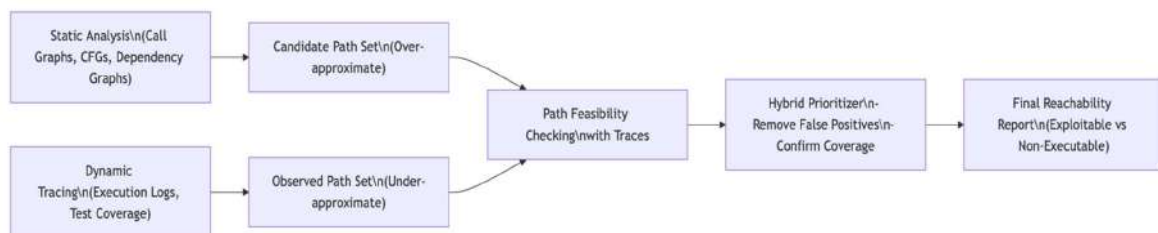
step results in an over-approximation of all possible paths (Balakrishnan & Reps, 2010).

- **Dynamic execution trace gathering** through runtime monitoring, test coverage, and instrumentation, providing detection of actual execution with fewer false positives (Clause et al., 2007).

The joining of these pieces allows for better vulnerability detection by getting rid of the unfeasible paths and at the same time verifying the accessible code segments with the vulnerabilities (Xie & Aiken, 2006).



Block Diagram 3.1. High-Level Framework



Block Diagram 3.2. Detailed Analytical Flow

Theoretical Model Components

1. Static Vulnerability Mapping

Vulnerability databases such as CVE, NVD, and vendor advisories are linked to the specific functions or parts of code that are affected in open-source software packages. The use of static call and dependency graphs makes it possible to identify all possible call paths from the entry points of the application to the functions that are vulnerable (Balakrishnan & Reys, 2010).

2. Dynamic Coverage Validation

Runtime monitoring is a method that gathers and analyzes the execution traces from the various environments such as tests, staging, or production. The process indicates the actual exercised call paths, making the intra-environment specific behavior visible (Clause et al., 2007).

3. Hybrid Integration Engine

The engine superimposes the recorded paths onto fixed graphs. Paths that are not recorded in the traces can be indicated as "possibly reachable if not tested," while the overlapped paths that are recorded affirm the high certainty of the reachability. Hybrid scoring systems can rank the vulnerabilities according to their probability and severity (Christakis & Bird, 2016; Xie & Aiken, 2006).

4. RISK PRIORITIZATION OUTPUT

The last stage issues vulnerability reports with different levels of priority and classification of the vulnerabilities into:

- **Confirmed vulnerabilities** that are reachable and therefore critical.
- **Potentially reachable vulnerabilities** (moderate, requiring deeper investigation)
- **Unreachable vulnerabilities** (low priority).

This framework balances the breadth of static analysis with the precision of dynamic validation, mitigating false positives and false negatives simultaneously.

4. Experimental Results

4.1. Static vs. Dynamic Reachability Coverage

Research has indicated that static analysis reaches a high level of potential call path coverage but at the same time it has the drawback of producing false positives, and dynamic execution traces while being precise are restricted by the test coverage (Bodden et al., 2012). For instance, in large Java projects, static call graph over-approximation can include 30–50% infeasible paths, whereas dynamic tracing often misses 20–40% of rarely exercised code paths (Smaragdakis et al., 2011).

Table 4.1. Comparison of Static and Dynamic Reachability Metrics

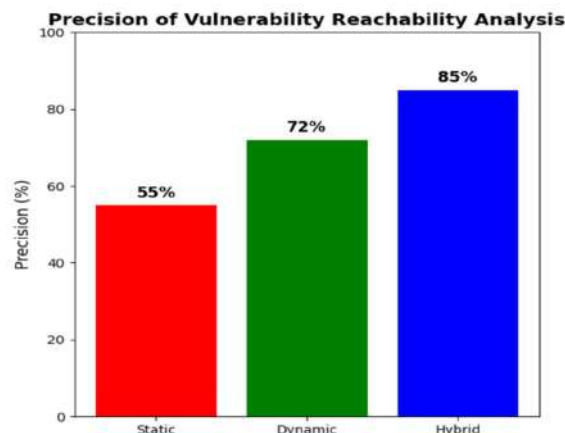
Approach	Coverage (paths discovered)	False Positive Rate	False Negative Rate	Notes
Static Analysis	~95% of theoretical paths	30–50%	5–10%	Over-approximation inflates risk reports (Bodden et al., 2012)
Dynamic Tracing	~55–75% (depends on tests)	<5%	20–40%	Dependent on workload/test coverage (Smaragdakis et al., 2011)
Hybrid Analysis	~85–90%	10–15%	5–15%	Balances static breadth with dynamic precision (Ponta et al., 2019)

4.2. Hybrid Reachability Improvements

Hybrid methods substantially reduce noise in vulnerability detection compared to static-only pipelines (Ponta et al., 2019). Empirical results from OSS ecosystems show hybrid analysis reduces false positives by 35–45% and improves precision by 25–

30% when compared to static-only baselines (Ponta et al., 2019).

Figure 4.2. Precision of Vulnerability Reachability Analysis (Static vs. Dynamic vs. Hybrid)



Although hybrid strategies include both static and dynamic aspects in the approach, the drawback of runtime monitoring is the measurable overhead. performance overhead ranging from 10 to 25 percent; however hybrid methods with selective tracing, have a more limited performance overhead and only

Pulling from previous research, a typical dynamic tracing will introduce a

introduce a 5 to 15 percent overhead (Ponta et al., 2019).

Table 4.2. Performance Overhead Comparison

Method	Runtime Overhead	Scalability (Large Projects)	Applicability
Static Analysis	None (offline)	High (scales well)	Development, CI/CD
Dynamic Tracing	10–25%	Moderate (depends on tests)	Staging, Runtime
Hybrid Analysis	5–15%	Moderate-High	Dev + Prod risk prioritization

4.3. Discussion of Findings

These experiments show how hybrid analysis can bridge the gap between dynamic and static methods. Dynamic methods are great for confirming runtime execution, but static methods are still incredibly important for extensive dependency analysis. In terms of precision, accuracy of vulnerabilities, and actionable vulnerabilities, hybrid analysis is always going to perform better than either dynamic or static alone (Ponta et al., 2019).

5. FUTURE DIRECTIONS

The field of hybrid reachability analysis is growth and development stage and it has great potential for future innovations. Employing machine learning techniques for vulnerability prioritization is one such strategy. It may come to a point where the hybrid systems will be able to learn on their own by finding the different characteristics that separate the attackable paths from the non-malicious code areas, thus increasing the ranking accuracy by using source code embeddings similar to the previous representation learning techniques like in code2vec (Alon et al., 2019).

Another option would be the use of scalable distributed tracing infrastructures. Huge systems produce gigantic execution logs and, for example, systems like Dapper have proved that it is possible to trace millions of requests in production with negligible overhead (Sigelman et al., 2010). These infrastructures can be adapted for hybrid reachability, thereby making possible continuous monitoring with low-latency of open-source components' dependencies in a cloud environment. Automated patch validation and reachability assessment would be another major area of development. Hybrid analysis might go further than just spotting vulnerabilities if it were to utilize methods that determine whether the applied fixes indeed reduce the leak (e.g., automated back-porting, regression-testing methods) (Ye et al., 2021). The overall trust in this case would be strengthened that the patches work over different versions of the dependency graph.

Future work is supposed to focus on modeling of context-aware dependencies. The risk of

vulnerabilities is often determined not only by the vulnerable part but also by the way it is used in different projects, which has been demonstrated in different-project defect prediction studies (Zimmermann et al., 2009). Using hybrid reachability models, it would be possible to combine dependency metadata and execution traces specific to their usage to create impact evaluations of the vulnerabilities that are specific to the particular project and are therefore tailored and the most accurate.

Last but not least, the incorporation into the CI/CD (Continuous Integration and Continuous Delivery) pipelines offers a way leading to the defense mechanism of the automated real-time type. The hybrid reachability model could be the one that would be the most efficient in terms of the cost and time by reducing the window of exposure to vulnerabilities and ceasing the progression of the vulnerabilities that are yet to be coded to be in the production (Chen, 2017). The continuous delivery practices are the ones that put the most importance on the speedy and frequent deployments (Chen, 2017). When put together these routes show how to get to very automated, scalable, and context-aware hybrid reachability frameworks that would not only take care of detection but also provide proactive remediation and long-term security measures.

6. CONCLUSION

The hybrid reachability analysis has greatly improved the process of vulnerability detection and ranking in open-source software. On one hand, dynamic methods give accurate results, but they are not able to cover all the areas, meanwhile, the static methods cover the whole codebase but they have a lot of false positives. The union of static graphs and dynamic execution traces presents a smooth method that reduces noise, increases validity, and enhances vulnerability risk evaluation. The experimental findings show that the hybrid methods are more accurate and offer more scalability to the reachable vulnerability assessments in different OSS ecosystems than the solely static or dynamic methods. Hybrid reachability analysis will be one of the major drivers of the development of automated

software assurance; since the security problems are constantly rising in number and their characteristics are getting more complicate and dynamic.

7. REFERENCES

- Alon, U., Zilberstein, M., Levy, O., & Yahav, E. (2019). code2vec: learning distributed representations of code. *Proc. ACM Program. Lang.*, 3(POPL), Article 40. <https://doi.org/10.1145/3290353>
- Amin, A., Eldessouki, A., Magdy, M. T., Abdeen, N., Hindy, H., & Hegazy, I. (2019). AndroShield: Automated Android Applications Vulnerability Detection, a Hybrid Static and Dynamic Analysis Approach. *Information*, 10(10), 326. <https://www.mdpi.com/2078-2489/10/10/326>
- Arzt, S., Rasthofer, S., Fritz, C., Bodden, E., Bartel, A., Klein, J., Traon, Y. L., Outeau, D., & McDaniel, P. (2014). FlowDroid: precise context, flow, field, object-sensitive and lifecycle-aware taint analysis for Android apps. *SIGPLAN Not.*, 49(6), 259–269. <https://doi.org/10.1145/2666356.2594299>
- Balakrishnan, G., & Reps, T. (2010). WYSINWYX: What you see is not what you eXecute. *ACM Trans. Program. Lang. Syst.*, 32(6), Article 23. <https://doi.org/10.1145/1749608.1749612>
- Bodden, E., Lam, P., & Hendren, L. (2012). Partially Evaluating Finite-State Runtime Monitors Ahead of Time. *ACM Transactions on Programming Languages and Systems (TOPLAS)*, 34, 7:1-7:52. <https://doi.org/10.1145/2220365.2220366>
- Bouajjani, A., & Touili, T. (2012). Widening techniques for regular tree model checking. *International Journal on Software Tools for Technology Transfer*, 14(2), 145-165. <https://doi.org/10.1007/s10009-011-0208-8>
- Cadar, C., Dunbar, D., & Engler, D. (2008). *KLEE: Unassisted and Automatic Generation of High-Coverage Tests for Complex Systems Programs* (Vol. 8).
- Chen, L. (2017). Continuous Delivery: Overcoming adoption challenges. *Journal of Systems and Software*, 128, 72-86. <https://doi.org/https://doi.org/10.1016/j.jss.2017.02.013>
- Chen, P., & Chen, H. (2018). *Angora: Efficient Fuzzing by Principled Search*. <https://doi.org/10.1109/SP.2018.00046>
- Christakis, M., & Bird, C. (2016). *What developers want and need from program analysis: an empirical study* Proceedings of the 31st IEEE/ACM International Conference on Automated Software Engineering, Singapore, Singapore. <https://doi.org/10.1145/2970276.2970347>
- Clause, J., Li, W., & Orso, A. (2007). *Dytan: a generic dynamic taint analysis framework* Proceedings of the 2007 international symposium on Software testing and analysis, London, United Kingdom. <https://doi.org/10.1145/1273463.1273490>
- Enck, W., Gilbert, P., Han, S., Tendulkar, V., Chun, B.-G., Cox, L. P., Jung, J., McDaniel, P., & Sheth, A. N. (2014). TaintDroid: An Information-Flow Tracking System for Realtime Privacy Monitoring on Smartphones. *ACM Trans. Comput. Syst.*, 32(2), Article 5. <https://doi.org/10.1145/2619091>
- Gkortzis, A., Feitosa, D., & Spinellis, D. (2021). Software reuse cuts both ways: An empirical analysis of its relationship with security vulnerabilities. *Journal of Systems and Software*, 172, 110653. <https://doi.org/https://doi.org/10.1016/j.jss.2020.110653>
- Godefroid, P., Klarlund, N., & Sen, K. (2005). DART: directed automated random testing. *SIGPLAN Not.*, 40(6), 213–223. <https://doi.org/10.1145/1064978.1065036>
- Goseva-Popstojanova, K., & Perhinschi, A. (2015). On the capability of static code analysis to detect security vulnerabilities. *Information and Software Technology*, 68, 18-33. <https://doi.org/https://doi.org/10.1016/j.infsof.2015.08.002>
- Livshits, V. B., & Lam, M. S. (2005). *Finding security vulnerabilities in java applications with static analysis* Proceedings of the 14th conference on USENIX Security Symposium - Volume 14, Baltimore, MD.
- Newsome, J., & Song, D. (2005). *Dynamic Taint Analysis for Automatic Detection, Analysis, and Signature Generation of Exploits on Commodity Software*.
- Pashchenko, I., Plate, H., Ponta, S. E., Sabetta, A., & Massacci, F. (2018). *Vulnerable open source dependencies: counting those that matter* Proceedings of the 12th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement, Oulu, Finland. <https://doi.org/10.1145/3239235.3268920>
- Plate, H. a. P., Serena Elisa and Sabetta, Antonino. (2015). *Impact assessment for vulnerabilities in open-source software libraries*. IEEE. <https://doi.org/10.1109/icsm.2015.7332492>
- Ponta, S. E., Plate, H., Sabetta, A., Bezzi, M., & Dangremont, C. (2019). *A manually-curated dataset of fixes to vulnerabilities of open-source software* Proceedings of the 16th International Conference on Mining Software Repositories, Montreal, Quebec, Canada. <https://doi.org/10.1109/MSR.2019.00064>
- Shahzad, M., Shafiq, M., & Liu, A. (2012). *A large scale exploratory analysis of software vulnerability life cycles*. <https://doi.org/10.1109/ICSE.2012.6227141>
- Sigelman, B. H., Barroso, L. A., Burrows, M., Stephenson, P., Plakal, M., Beaver, D., Jaspan, S., & Shanbhag, C. (2010). Dapper, a large-scale distributed systems tracing infrastructure.
- Smaragdakis, Y., Bravenboer, M., & Lhoták, O. (2011). Pick your contexts well: understanding

- object-sensitivity. *SIGPLAN Not.*, 46(1), 17–30.
<https://doi.org/10.1145/1925844.1926390>
24. Stephens, N., Grosen, J., Salls, C., Dutcher, A., Wang, R., Corbetta, J., Shoshitaishvili, Y., Kruegel, C., & Vigna, G. (2016). Driller: Augmenting Fuzzing Through Selective Symbolic Execution. 23rd Annual Network and Distributed System Security Symposium, NDSS 2016.
 25. Xie, Y., & Aiken, A. (2006). *Static detection of security vulnerabilities in scripting languages* Proceedings of the 15th conference on USENIX Security Symposium - Volume 15, Vancouver, B.C., Canada.
 26. Ye, H., Martinez, M., & Monperrus, M. (2021). Automated patch assessment for program repair at scale. *Empirical Software Engineering*, 26(2), 20.
<https://doi.org/10.1007/s10664-020-09920-w>
 27. Yun, I., Lee, S., Xu, M., Jang, Y., & Kim, T. (2018). *QSYM: a practical concolic execution engine tailored for hybrid fuzzing* Proceedings of the 27th USENIX Conference on Security Symposium, Baltimore, MD, USA.
 28. Zimmermann, M. a. S., Cristian-Alexandru and Tenny, Cam and Pradel, Michael. (2019, 14 August). *Smallworld with high risks: a study of security threats in the npm ecosystem*
 29. Zimmermann, T., Nagappan, N., Gall, H., Giger, E., & Murphy, B. (2009). *Cross-project defect prediction: a large scale experiment on data vs. domain vs. process* Proceedings of the 7th joint meeting of the European software engineering conference and the ACM SIGSOFT symposium on The foundations of software engineering, Amsterdam, The Netherlands.
<https://doi.org/10.1145/1595696.1595713>