

CSEC 793 CAPSTONE IN COMPUTING SECURITY
PROJECT REPORT

MS CAPSTONE REPORT

April 2, 2018

Ryan Whittier
Department of Computing Security
College of Computing and Information Sciences
Rochester Institute of Technology
rjw4910@g.rit.edu

1 Introduction

The project in this paper is the development of a security focused continuous integration (CI) pipeline. The goal is to make a pipeline that focuses on security in an attempt to improve web application security. This pipeline will automatically run a series of tests against a repository of configuration files and code. These tests will enforce correct configurations and security practices within committed changes.

CI is a technique that is used to lead to better development practices, but also to speed up development releases. Those benefits can also be tailored to improve security in development. Many penetration tests or bug reports revolve around the same information repeated for different cases. These reports often have a common solution. If we discuss web applications, one common vulnerability is cross site scripting (XSS). XSS vulnerabilities have the short term recommendation of encoding all user input to avoid users input being interpreted as part of the webpage. The long term recommendation is to set up a full Content Security Policy (CSP), limiting where scripts and HTML tags can be loaded from. Companies generally choose the short term route of fixing each individual case, but this will cost them time and money in the long term, eventually they decide to implement a long term solution. In this example companies will often implement encoding, but as they expand the web application, they find the same bug. The company eventually switches to creating a CSP and configuring it correctly so new expansions can not be affected by a XSS bug.

Everyday applications are getting larger and are handling enormous amounts of data. Many of these applications are services that companies provide, such as Facebook and Netflix. All of these applications need to provide security for both their internal company data and the end user's data. The problem is that as these applications grow, their logic gets more complex and mistakes will be made, potentially introducing vulnerabilities into the codebase without the developers knowing it. Sometimes these mistakes are simply forgotten flags on cookies and other times they can be complex flaws in the business logic of the application. These mistakes can result in damage ranging from defacement of a company site, to a complete breach of a company's internal network.

A bug does not need to be exploited by a malicious actor to cause financial loss for the company. Many companies offer bug bounty programs, in which a researcher can discover and report a bug for a reward. The rewards vary by company, but often times the researcher will receive a substantial monetary reward, depending on the severity of the bug discovered. This means that each time the company releases updates to their applications, researchers will be scavenging for more vulnerabilities, which can costly for the company. It may also cost the company's security team a significant amount of time, as they need to look into and verify each report that is submitted. Often times, researchers can find very common vulnerabilities in the application, that usually result from developers attempting to push out new features at such a rapid pace.

CI can be used to catch simple mistakes automatically, by ensuring that the common bugs are found before the changes are deployed into production. By finding these bugs

before they are introduced into production, the number of submissions from bug bounty programs will decrease dramatically. This means that the bug reports that are submitted by researchers will be of higher quality and provide more value to the company.

2 Literature Review

2.1 Introduction

Continuous Integration (CI) has grown in popularity, being used in both enterprise and open source projects. CI provides an increase in productivity to programming projects by creating an environment where building, testing, and often deployment is done automatically. When developers do not need to do these tasks manually they can spend more time programming and finding bugs that trigger a failed build or failed test. Most research into CI looks at increasing productivity of developers and the enforcement of development standards. The potential use of CI for security has been mostly overlooked with most research looking at exploiting and protecting a CI system or a CI pipeline and not at the benefits of a pipeline focusing on application security.

A separate pipeline for security could provide a number of benefits for a security baseline. It could force certain configurations such as Content Security Policy in web applications, or lack of use of depreciated crypto algorithms such as DES or TripleDES [14]. A pipeline that looks only at big win configurations such as these would serve a similar function as Automated Source Code Analysis Tools (ASCAT) do when looking at code meeting project coding guidelines [15]. The pipeline could also include ASCATs that search for security bugs in a warning stage to avoid wasted time with failed builds from false positive. Another potential feature is an inclusion of fuzzers, which were recently used against memory forensic tools to look into anti-forensics techniques through crashing the tools [3].

2.2 Test case generation

Test case generation is a discipline that focuses around automatically generate unit tests for Software Engineering projects. The discipline sometimes focuses on generating tests from a base, such as the source code or a config file, and generating a set of tests from the base or it focuses on taking existing tests and exanpding them to cover more features.

One paper from the 2014 Automated Software Engineering conference looked at generating test cases for web from an existing selenium test suite [4]. This paper goes through the method that is used to generate the new test cases. The method starts by taking a selenium test suite and going though it keeping track of states, http element interactions, and test case assertions. It uses the base states and element interactions to create a State-Flow Graph, which is a graph of the states with the interations that that move from one state to the next and the assertions that are involved to a single state. It then uses a web crawler on each state to discover alternate paths and new states. It then makes use of assertions on the base states to generate new tests for the crawler discovered states. An example given in the paper is that an application that allows you to create notes is tested with selenium by logging in, clicking the create note element, and then verifying that an id element had specific text. The test suite was then expanded by crawling to find the edit note interaction and the delete note interaction. These were then clicked and tested using the assertion,

verifying that the id was present and changing the text to match the new interactions result text.

2.3 Test case selection

The idea behind test case selection is limiting tests run to save on time during builds. As a code base grows, the number of tests increase. The increase in tests make it more and more expensive and eventually becomes to big a load to do everytime [12, 13]. This has lead to ways of avoiding running all of the tests on every code change. The ways of running tests have historically been either rerun all of the tests or manually pick test cases to run based off of the changes the developer made.

A paper by Milos Gligoric looks at comparing manual test selection against research made into automatic test selection, where automatic test selection would only run the tests that test code affected by the changes a developer made [7]. The state of the automatic test selection tools were limited to Google and Microsoft at the time of the paper. The tools were impractical for use by smaller projects because they were either too imprecise, selected tests that were not affected by the code changes, or unsafe, failed to guarentee all tests not selected were unaffected by the code changes. The paper specifically compares 14 developers manual test choices against a tools automated test selection. The paper found that manual test case selection was most commonly done during debugging. It also found the 73% of the time manual RTS selected more tests than were necessary and 74% of the time some tests that were affected were missed.

2.4 Effective use of CI

Continuous integration has been a fantastic addition to Software Engineering practices by removing repetitive administrative tasks. The less steps a software engineer has to go through when producing new code, the more time they can spend on adding features, squashing bugs, or otherwise improving the project code base. There is plenty of research done into the use of CI including many case studies, papers and presentations looking at how to setup CI and Continuous Deployment (CD), and papers looking at specific configurations of CI.

CI can be integrated with all sorts of tools to increase productivity. The paper How Open Source Projects use Static Code Analysis Tools in Continuous Integration Pipelines by Fiorella Zampett looks at how Static Code Analysis tools in a CI pipeline effects projects [15]. The paper found that the most utilized feature of ASCATs was to ensure a project's code was consistent by ensuring the developers coding guidelines were met. The paper also found that ASCATs caused broken builds to be fixed quickly in an average of 8 hours and one build. There are a number of recommendations that this paper provides which outline how to set up ASCATs in a CI pipeline, what to think about when doing so, and what to expect to maintain the ASCAT. Adding a source code analysis tool will help keep a project

consistent while helping point out bugs that could be missed by developer made unit tests.

Two papers look at CI and how it affects projects in general. Continuous Integration and Quality Assurance: A Case Study of Two Open Source Projects by Jesper Holck and Usage, Costs, and Benefits of Continuous Integration in Open-Source Projects by Michael Hilton both look at open source projects and how CI affects them [8, 9]. The papers found that CI often replaces the practice of having developers make formal design documents. The developers instead just pick from a list of tasks and works on completing them and merging them into the code base. The papers also found that most developers like CI and plan to use it again in future projects. For projects that did not use CI it was found that the reason was usually just that the developers were not familiar with how to set up and use CI. Both papers found CI to be successful in increasing productivity, causing projects to release twice as often, accept pull requests quicker, and have developers less worried about breaking the project.

2.5 Security in CI pipelines

Security conferences often look into how to break systems as a way of pointing out the flaws of a configuration. CI servers have also had their share of security professionals testing for exploits and looking at the consequences of compromise. CI has also had a little bit of research into how to secure a CI pipeline.

A presentation at DEFCON named Exploiting Continuous Integration (CI) and Automated Build Systems talked about the consequences of a CI enabled projects [11]. The presentation found that exploiting a repo that holds a CI integration ends with a huge amount of access. If the repo links into an internal CI server, then the attacker ends up with internal network access. If the repo links to a CI server with multiple CI instances or also runs the CD, then the attacker will get more source code and access to the deployment machines because the CI holds a way to connect to the deployment servers. Otherwise the attacker ends up with environment variables which often hold extremely sensitive information.

A paper by Len Bass called Securing a Deployment Pipeline looks at how to secure a CI pipeline to limit the damage in case of exploitation [2]. The paper details a way to break a pipeline down into trusted and untrustworthy parts, segmenting operations until an untrustworthy segment cannot be broken down any more. The CI pipeline then holds parts that are guaranteed to be trustworthy and run as expected, minus specific cases outlined in the paper, and parts that may provide untrustworthy output. By limiting the scope of untrustworthy parts, the rest of the pipeline can run as expected and it is possible to see where the most risk lies. Then the owners of the pipeline can work to limit access to the untrustworthy portions and look into solutions to make those portions trustworthy.

2.6 Use of CI for Security

There is little research into the use of CI for security. One presentation by Mozilla looks at the use of CI to tackle easy fixes in response to their bug bounty program [14]. The presentation looks at using CI to ensure that the production environment contains configurations that mirror best practices for common web application security bugs. Some examples include HSTS is enabled, CSP for XSS bugs, various X-Options headers, Cookies have secure, Cross origin sharing, and Subresource integrity checks. The presentation recommends figuring out a security baseline for a projects CI pipeline, drive testing from the CI pipeline, and empower the team to fix the issues. Another recommendation is not to break on deployment into production as that could break the production site if configured poorly. The end result of mozilla's CI setup was a large drop in bug reports that the CI tests aimed to fix.

2.7 Work and methodology

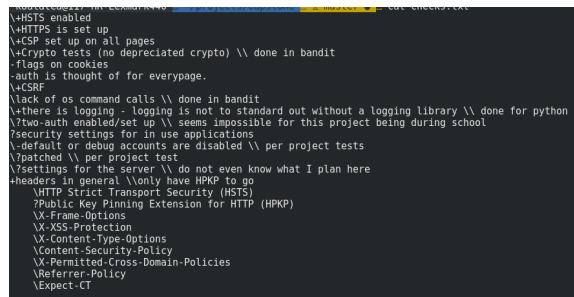
The work into CI so far has shown that CI is useful tool for developers to ease the creation of new features. CI has spread to open source projects and is deployed in most organizations that have at least one large code project. The security side of how a CI is dangerous if exploited and how to secure a pipeline against attacks has had little research. The most interesting thing that I think is lacking is the look into how CI can be used to improve the security the project that it is integrated on.

Most of the research does not mention how CI could help improve the security of a project. Some ways this could be done are through targeting common bugs that are already fixed. Some examples are ensuring that CSP is enabled, HSTS is enabled, parameterized queries are used, binary protections are enabled in compilation scripts, and unsafe functions are not used. Some of these are implemented already in some source code analysis tools, but these tools often have high levels of false positives. Another use could be in including fuzzing in a pipeline. I have not seen any research into automating fuzzing into testing an application. Depending on the project it could be a very useful tool to find bugs.

The research I am doing involves creating a pipeline with tests that focus on specific frameworks and are tailored towards an application instead of general language based tests. Most source code analysis tools look at specific languages instead of focusing on specific libraries or frameworks [1, 6, 5]. My tests will also do some dynamic testing instead of just focusing on source code analysis to verify that the protections are in place in a live application. I am following a similiar path to Mozilla in using tests to mandate security configurations [14].

3 Project Implementation

The projects goal is a way to mandate security with low false positives. The start of this project involved creating a list of checks to implement as shown in Figure 1. The checks



```

+HSTS enabled
+HTTPS is set up
+CSP set up on all pages
+Crypto tests (no deprecated crypto) \\ done in bandit
-flags on cookies
-auth is thought of for everypage.
+CSRF
\\lack of os command calls \\ done in bandit
+there is logging - logging is not to standard out without a logging library \\ done for python
?two-auth enabled/set up \\ seems impossible for this project being during school
?security settings for in use applications
\\default or debug accounts are disabled \\ per project tests
?patched \\ per project test
?settings for the server \\ do not even know what I plan here
+headers in general \\only have HPKP to go
  \\HTTP Strict Transport Security (HSTS)
  ?Public Key Pinning Extension for HTTP (HPKP)
  \\X-Frame-Options
  \\X-XSS-Protection
  \\X-Content-Type-Options
  \\Content-Security-Policy
  \\X-Permitted-Cross-Domain-Policies
  \\Referrer-Policy
  \\Expect-CT

```

Figure 1:
The list of checks to implement

were chosen for both their ease of being checked and their security relevance. Tests that could be dynamic were prioritized over those that had to be static. The notation in Figure 1 checks is that a leading \ means the test is addressed, a + means that the test has been chosen to be implmented, a - meant the test would not be implemented, and a ? meant that their are still questions about the test. The majority of tests implemented are headers, with each one making specfic bugs harder to exploit.

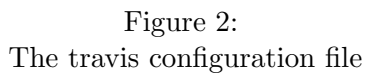
The implemented tests were all of the headers tests, and the logging test. The depreciated crypto test and the os test were both found to be done for python in the bandit source code analysis tools [1]. The HPKP header was dropped due to it's pending removal from chrome in favor of the new Expect-CT header [10]. The headers are chosen because they are easy to test for dynamically, the same rational applies to csrf and https. The only static test that is created for this project is the logging test one because it does not require context like authentication being addressed for every web page.

The HTTPS test is tested by verifying that browsing to HTTP root redirects to an HTTPS root and then that doing a GET request against the HTTPS root returns a 200 response. Following the HTTPS test, HSTS is tested next by checking the Strict-Transport-Security header and that at least max-age is set, further parts could be included if the developer also wanted includeSubDomains or preload. CSP is tested in a similiar way by checking that the Content-Security-Policy header is present and does not have unsafe-inline within any of the Content-Security-Policy headers. The rest of the headers are implemented soley by checking for their existence. The headers have settings that are too different between projects so they are not tested for. The CSRF test is implemented using a list of dictionaries to keep track of end points. Each dictionary is of an endpoint that has a form to fill. The dictionaries have the endpoint, data to post, fail_text which is the text present if the form

submission failed, `success_text` which is the text present if the form submission succeeds, a `fail_status_code` which is the status code for a failed submission, and a `success_status_code` if the submission succeeds. In the case of there not being one of the four success or fail codes or text then they are set to none. In the CSRF test for each endpoint the endpoint is queried and is searched for an input html tag with an id of `csrf_token`. If the `csrf_token` exists the endpoint is POSTed to without the token. The response to this post is checked for each fail condition that is not none, and reach success condition that is not none is checked to make sure they are absent. Finally the `csrf_token` found earlier is added to the data and POSTed again with the checks from the POST without the token are reversed. The rest of the tests are static tests. Only one was implemented because it was found that the rest of the planned ones were already implemented. The os tests are done for potential inject points using os, which is slightly different from the original goal in the checks list, but the crypto one does exactly what is being looked for by checking for depreciated crypto. The written test was written as a plugin for bandit, it checks for any print statements and recommends using a logging library instead. This one exists partially because most developers already use logging libraries for most projects, the other reason is for proper logging management to avoid resource exhaustion.

All of the dynamic tests were easy to transfer over to java, only the url and the requests library had to be changed and the test functioned that same. The caveat is that the java application had to be dockerized so that it could be tested with the python code, while the python tests worked by using flasks built in debug client. The static tests are not as easy to transfer over. To use the static test the already implemented tests either needed to be found in another static code analysis tool or written again for java.

These tests were then added to a CI pipeline in Travis. To do so build stages were used for parallelizing the tests. In this project this made sense because there were two passing and two failing builds. Figure 2 is the `travis.yml` configuration file.



4 Conclusions

In this section, you should provide a concise summary of your work, state the importance of your idea and your contribution towards solving the problem. Point out any improvement could be done if you had more time. List some ideas as future work.

References

- [1] *Bandit python source code analysis tool*. URL: <https://github.com/openstack/bandit>.
- [2] Len Bass. “Securing a Deployment Pipeline”. In: *IEEE/ACM 3rd International Workshop on Release Engineering* (2015), pp. 4–7.
- [3] Andrew Case. “Gaslight: A comprehensive fuzzing architecture for memory forensics frameworks”. In: *DFRWS USA d Proceedings of the Seventeenth Annual DFRWS USA* (2017), S86–S93.
- [4] Amin Milani Fard. “Leveraging Existing Tests in Automated Test Generation for Web Applications”. In: *29th IEEE/ACM International Conference on Automated Software Engineering (ASE)* (2014), pp. 67–78.
- [5] *Find Security Bugs java source code analysis tool*. URL: <https://find-sec-bugs.github.io/>.
- [6] *FindBugs java source code analysis tool*. URL: <http://findbugs.sourceforge.net/>.
- [7] Milos Gligoric. “An Empirical Evaluation and Comparison of Manual and Automated Test Selection”. In: *29th IEEE/ACM International Conference on Automated Software Engineering (ASE)* (2014), pp. 361–372.
- [8] Michael Hilton. “Usage, Costs, and Benefits of Continuous Integration in Open-Source Projects”. In: *International Conference on Automated Software Engineering (ASE)*. (2016), pp. 426–437.
- [9] Jesper Holck. “Continuous Integration and Quality Assurance: a case study of two open source projects”. In: *Australasian Journal of Information Systems 11(1)* (2012), pp. 40–53.
- [10] *Intent To Deprecate And Remove: Public Key Pinning*. URL: <https://groups.google.com/a/chromium.org/forum/#!msg/blink-dev/he9tr7p3rZ8/eNMwKPmUBAAJ>.
- [11] spaceb0x. *Exploiting Continous Integration (CI) and Automated Build Systems, DEF-Con 25*. URL: <https://media.defcon.org/DEF\%20CON\%2025/DEF\%20CON\%2025\%20presentations/DEFCON-25-spaceB0x-Exploiting-Continuous-Integration-UPDATED.pdf>.
- [12] *Testing at the speed and scale of Google, 2011*. URL: <http://goo.gl/OKqBk>.
- [13] *Tools for continuous integration at Google scale, 2011*. URL: <http://www.youtube.com/watch?v=b52aXZ2yi08>.
- [14] Julien Vehent. *Test Driven Security in Coninuous Integration*. Presentation. Enigma. 2017. URL: <https://www.usenix.org/conference/enigma2017/conference-program/presentation/vehent>.

- [15] Fiorella Zampetti. “How Open Source Projects use Static Code Analysis Tools in Continuous Integration Pipelines”. In: *IEEE/ACM 14th International Conference on Mining Software Repositories (MSR)* (2017), pp. 334–344.