Software Engineering Assignment 1
WASP Course 2024

Mojtaba Moazen

September 3, 2024

1 introduce to the Software Supply Chains

My research focuses on enhancing software security within the software supply chain, particularly emphasizing integrating security practices into the Continuous Integration and Continuous Deployment (CI/CD) pipeline. Different types of attacks, which exploit vulnerabilities in third-party dependencies, development processes, and the tools used throughout software development and deployment, are increasingly targeting the software supply chain. During the past years of development, many tools have been developed in order to implement different aspects of CI/CD in software development including develop, test and deploy. Therefore, the security of these tools is becoming important to industry and we need robust and reliable methods in this era. Without a secure and reliable method, CI/CD pipelines can be exploited through vulnerabilities and much research confirms this [2].

Right now, my primary focus is on identifying issues in CI/CD tools, and more crucially, addressing this question: How can we secure a particular pipeline against a specific attacker model? In this research area, we aim to ensure that security is not an afterthought but an integral part of the development process. This process includes static and dynamic code analysis, dependency scanning and secret management in the production environment.

Another research area we are focusing on is evaluating the effectiveness of security tools that have been developed aim to identify vulnerabilities and unauthorized access during the CI/CD pipelines. By this, software development can shift security left and address potential problems and issues before the software deploys on production. mane research have been done

regarding this during the past years of study. This process contains validation of verification methods in software supply chains aim to make sure that methods are working correctly. By systematically evaluating and refining these security tools, our goal is to ensure that organizations can confidently implement security measures that are both robust and efficient. This process help to organization to manage the security risks in a better way and maintenance of software supply chains method would be easier.

2 Ideas from Robert's lectures

I will first focus on the first idea that I got from Robert's lecture: **Behavioral Software Engineering: People are not rational**. This concept, Behavioral Software Engineering, has direct implications for the security of CI/CD pipelines and therefore software supply chains. As we discussed in the lecture, People are not rational always based on this concept, therefore, engineers, developers and even DevOps engineers are prone to cognitive bias or irrational behavior. Even with respect to security best practices for automated tools, developers may introduce vulnerabilities in the process of software development specifically CI/CD pipelines. Consequently, this shows the importance of security check methods on pipelines. This brings the question is whether these methods are sufficiently secure. Another concern related to this concept is that, developers often prioritize speed and functionality over security. This explain the reason for an organization which prioritize the faster delivery to slower security checks. another reason that we need a secure and reliable automated checks on CI/CD processes.

As the second idea that I understood during the lecture, was the difference between **Validation and Verification**. In the concept of software supply chains, verification ensures that each component or dependency, whether developed in-house or sourced from third-party vendors, meets its design specifications. This is importance to software supply chains since we use open source component and modules during the software development which may introduce vulnerabilities if not properly verified.

On the other hand, validation in the software supply chain ensures that the entire system works based on design, after integrating all components. This step is vital because even if individual components are verified, the system as a whole must be validated to confirm that it behaves securely and correctly when deployed. This become importance when we use a third-party component. Through the CI/CD framework, validation can be happen by testing and automated deployment process.

3 Ideas from guest lectures

In lecture we had a guest lecture from VolVo and discussed about SE4ML and ML4SE concepts. During the lecture we had an example of how VolVo applied these concepts on their production. As we discussed, SE4ML focuses on applying software engineering principles to the development, deployment, and maintenance of machine learning systems. Just like a traditional software, machine learning models need to be deployed as an application and need automation with respect to the develop, text and deploy stages. As the result, we need to make pipeline secure and reliable ensuring that the models are not tampered with during deployment and that they perform securely in production environments aligns with your focus on securing the CI/CD process. ML softwares are based on third-party libraries and this can result in vulnerabilities exploit [1]. On the other hand, we have discussed about the ML4SE, related to use of machine learning to detect different vulnerabilities and problems in supply chains. Machine learning models trained on large code bases can help assess code quality and predict security risks. By integrating these models into the CI/CD pipeline, you can automate the identification of potential security flaws before they make it into production and this definitely help to detection process.

4 Related paper to the Software Supply Chains from CAIN Conference

I found 2 more related paper from CAIN 2022 and 2023. In the following I wil provide a details of how these paper are related to my research with focusing on integrating and provide complementary solution for software supply chains analysis.

As the first paper , I found this paper related to software supply chains [4]. This paper presents an approach for assessing security risks in ML systems, which can be integrated into your research on CI/CD pipelines. Proposed method involves a series of questions with Yes or No answer based on the attack tree. This research is critical when we need to integrate the security assessment into CI/CD pipelines specifically when we are deploying a ML based application or model. Integrating this approach with CI/CD pipelines can help to detect the security risk regarding the ML model and help to detect the vulnerabilities in this area.

As the second paper, I found this paper interesting [3]. They suggested Tenet, a modular approach based on machine learning to find vulnerabilities

in softwares. It emphasizing on integrating the detection approach with CI/CD pipeline to make it more automate. Eventually we always need to work on the security guarantees from these approach by analysis and validate the method they used and this is directly connected to the software supply chains and related research area. This can be done through validation and verification of supply chains methods.

References

[1] M. S. Akter, M. J. H. Faruk, N. Anjum, M. Masum, H. Shahriar, N. Sakib, A. Rahman, F. Wu, and A. Cuzzocrea. Software supply chain vulnerabilities detection in source code: Performance comparison between traditional and quantum machine learning algorithms. In *2022 IEEE International Conference on Big Data (Big Data)*, pages 5639–5645, 2022. doi: 10.1109/BigData55660.2022.10020813.

[2] I. Koishybayev, A. Nahapetyan, R. Zachariah, S. Muralee, B. Reaves, A. Kapravelos, and A. Machiry. Characterizing the security of github CI workflows. In *31st USENIX Security Symposium (USENIX Security 22)*, pages 2747–2763, Boston, MA, Aug. 2022. USENIX Association. ISBN 978-1-939133-31-1. URL `https://www.usenix.org/conference/usenixsecurity22/presentation/koishybayev`.

[3] E. Pinconschi, S. Reis, C. Zhang, R. Abreu, H. Erdogmus, C. S. Păsăreanu, and L. Jia. Tenet: A flexible framework for machine-learning-based vulnerability detection. In *2023 IEEE/ACM 2nd International Conference on AI Engineering – Software Engineering for AI (CAIN)*, pages 102–103, 2023. doi: 10.1109/CAIN58948.2023.00026.

[4] J. Yajima, M. Inui, T. Oikawa, F. Kasahara, I. Morikawa, and N. Yoshioka. A new approach for machine learning security risk assessment – work in progress. In *2022 IEEE/ACM 1st International Conference on AI Engineering – Software Engineering for AI (CAIN)*, pages 52–53, 2022. doi: 10.1145/3522664.3528613.