

Detecting Unsafe Updates in Software Ecosystems

Yao-Wen Chang
Student ID: 1346258

Supervisor: Christoph Treude
Industry Mentor: Behnaz Hassanshahi (Oracle Labs, Australia)

Word Count: X words

Submission Date: September 17, 2023

Contents

1	INTRODUCTION	2
2	RELATED WORK	2
2.1	Definition of CI/CD	3
2.2	Three Main Stage in CI/CD	3
2.2.1	Source Code Repository	3
2.2.2	Build/Test	3
2.2.3	Deploy	4
2.3	OWASP Top 10 CI/CD Risks[5]	4
2.4	U.S. Department of Defense Recommend [12]	7
2.5	in-toto Framework	7
2.6	Conclusion	8
3	BACKGROUND	8
3.1	What is Software Supply Chain?	8
3.2	Provenance	8
3.3	SLSA Provenance versus SBOM (Software Bill of Materials)	8
3.4	Build Model	9
3.5	SLSA Security Level	9
3.6	Policy Engines	9
4	METHODOLOGY	10
4.1	Research Aims	10
4.2	Research Objectives	11
5	TIMETABLE AND PLAN	11
5.1	Phase One	11
5.2	Phase Two	11
5.3	Phase Three	12
5.4	Timeline	13
6	CONCLUSION AND EXPECTED OUTCOMES	13
6.1	Conclusion	13
6.2	Expected Outcomes	13

ABSTRACT

This literature review will introduce Continuous Integration/Continuous Delivery (CI/CD). Also, some attack surfaces and how the malicious attackers exploit these attack surfaces to compromise CI/CD pipelines will be covered in this review. Different methods and frameworks are going to be introduced to counter the attack. Some method target risks at certain stage in the pipeline, but some cover the whole pipeline.

In our research, the framework will focus on Supply Chain Level Security Artifacts (SLSA) which is adopted by Google.

1 INTRODUCTION

CI/CD is a development method used to efficiently build and test code updates. It helps organizations keep their software consistent and allows them to smoothly incorporate new changes. However, CI/CD systems can be tempting targets for cyberattackers. These attackers may try to insert malicious code into CI/CD processes, steal valuable credentials, or disrupt the original functioning of applications.

Recent incidence like the infection of SolarWind’s Orion platform [8, 14] which is used to monitor and manage the network is downloaded by thousands of customers, including U.S. government agencies, critical infrastructure providers, and private companies.

Another malicious attack target the credentials from a contributor of the esline-scope package. The attackers updated the malicious code within the code base, which will end up stealing multiple credentials from the downstream users [3].

Sometimes, the risks are due to the situation that the consumers are not aware of the importance of frequent update of the dependencies which are consumed in their packages. In this situation, the attackers is able to discover the unsafe dependencies on the git repositories, then looking for other packages depend on these unsafe outdated dependencies. It is simple to find out how the bugs being fixed on the original dependencies’ repositories, and directly compromise the downstream users with the bugs. Therefore, a smart alert system [17] to detect if the older version contains bugs and should be updated. Then, the system will automatically update the dependencies. This system can be deployed in different stages to automatically detect unsafe version, and update the dependencies. The developers are sometimes not aware of using the older dependencies will result in tremendous attack. With this automatic tool, it will support developers to solve the issues.

In Section 2, we delve into the related work on the CI/CD pipeline security. Three main attack surfaces within CI/CD pipeline and the counter measures are introduced here. Additionally, we will summarise the OWASP Top 10 CI/CD Risks. These will provide a comprehensive foundation for our research.

In Section 3, we will explore the realm of Software Supply Chain Security, which is the prerequisite for understanding the key components of the Macaron Framework implemented in our research. We dissect the Software Supply Chain, emphasizing the critical importance of provenance. Also, we will explain the nuanced difference between Software Bill of Materials (SBOM) and the SLSA Provenance. Finally, we investigate the intricacies of the Build Model and the Security Levels specified by the SLSA framework.

In Section 4, we delineate our research aims and objectives. Our overarching aim is to fortify the security aspects of CI/CD pipelines. And our objectives include how we are going to achieve our aims. Furthermore, the current research progress and the research gap we find out in related works will be improved by our contribution.

Section 5 provides a detailed timetable and plan for the research project. We outline the key milestones, tasks, and deadlines necessary for successful project completion.

In Section 6, we offer a concise conclusion to the research proposal, summarizing our intentions and expectations.

2 RELATED WORK

In the literature review, first, prerequisite definition of the CI/CD technique will be described. Then, a variety of risks and attacks are categorised based on the three attack surface within the CI/CD pipeline, which are source code repository, build and test, and deploy. Some defense methods introduced in the related work will be aligned with the attacks in this section. Finally, the top ten risks existed in CI/CD pipeline referring to OWASP 2023 will be summarised [5].

2.1 Definition of CI/CD

CI/CD is a development and deployment process for automatically building and testing code changes that support organizations maintain a consistent code base, then automatically shipping the artifact to the target customers. 'CI' involves developers frequently merging code changes into a central repository where automated builds and tests run. 'Build' is the process of converting source code into executable code. Then, running the automated tests which usually combined unit test and integration test against the build. These process will avoid integration challenges that can happen when waiting for release day to merge changes into the release branch. 'CD' defined the process of the releases happened automatically [12].

Nowadays, nearly all software applications were built on top of others works, which is called third-parties. However, the convenience and capabilities of the third-party source code usually cause cybersecurity risks [10]. Software supply chain attacks aim at injecting code into software components including source code base and artifacts, like docker image or executable software, to compromise downstream users [8]. Some risks and counter method will be introduced in the following sections.

2.2 Three Main Stage in CI/CD

In the following three subsections, we will categorise some attack events and potential attack methods into the high-level representation of three main categories within the CI/CD pipeline. This categorisation allows us to gain a deeper understanding of the security challenges and risks associated with each phase of the software development and delivery process.

2.2.1 Source Code Repository

Java Virtual Machine (JVM) executes Java bytecode and provides strong safety guarantees. However, the unsafe API, "sun.misc.Unsafe", will cause serious security issue if it is misused by the developers. The research [10] studied a large repository, Maven, and analyzed the compiled Java code. The security issues include violating type safety, crashing the virtual machine (VM), uninitialized objects and so on. These misuse might impact third-party package management service.

The research[6] proposed anomaly detection method to cluster and detect the suspicious update JavaScript code. It shows that this unsupervised machine learning clustering method is able to reduce the manual review effort by 89 percent considering the worse case that all updates are not malicious. However, the data is skewed since more unsuspicious updates existed in the dataset. Also, the researcher divide the features based on static method by reviewing some packages and find out the malicious code. This method is not cost-efficient and only specific to the programming languages where there are models already trained.

Nevertheless, the author in [2] provide an attack approach to bypass the manual code review. With the Trojan source attacks, the attacker can embed seemingly unsuspicious Unicode into the comments and indirectly modify the order of the code causing unexpected results. This would bypass the method provided by [6], since this feature is not recognised by the model. The author recommends the developers to include Regex detection method in the 'Build' stage, since this malicious code is triggered by the compiled code. The fundamental way to avoid being compromised by the Trojan source attack is through banning the Unicode when it is not required in the projects. Even though recommend defense strategies are able to counteract this attack, but the method only deal with the specific attack against the CI/CD pipeline.

2.2.2 Build/Test

Malicious code injection can occur during the build of the pre-built components, especially in compiled languages. This attack method is favored by the attackers, since the detection of the pre-built components and compiled code is typically more difficult [9]. Therefore, opting for building package directly from source code is a measure to avoid this risk.

In [13], an overview of the attack methods and how the attacks are triggered are being summarised. The most frequent way implemented to inject the bad dependencies is 'Typosquatting'. It is a simple method to modify the packages' name with a similar one. If the consumers did not carefully check the installed packages, the malicious dependencies will being triggered based on the trigger point designed by the attackers. Also, this research find out the popular trigger point of the malicious dependencies is at the installation step. At the installation step, the malicious code is usually embedded in the installation

script. When the customers install the packages, the systems are directly being compromised. This research paper implements the taxonomy method to deduce possible trigger points and the possible attacks occurring within each CI/CD stage. This method provides developers a clear and completed checklist to prevent the potential vulnerabilities. However, this paper does not provide the completed counter measure against the attacks.

2.2.3 Deploy

In the extremely severe attack event [18], the unknown entity gain access to the GitHub repository with the higher permission from the controlled account of the repository maintainer. The accident causes malicious code to be committed and the attacker even trying to remove various repositories. Some good practices are implemented in this repository. For example, the audit logs function provided by GitHub enables the Gentoo organization to respond quickly to prevent from the extra impact. However, the Gentoo GitHub organization does not implement 2FA to prevent the accounts from being fully compromised. Also, it does not manage the identities correctly.

[16] presents a framework to detect the discrepancies of the packages on between source repositories and package managers. This framework can be cooperated with other tools to improve efficient and effective. However, the research does not consider the situation that there are no discrepancies between the source code and the published packages, but the code have already been maliciously injected. In the framework we deploy in this research, the provenance can ensure there are no suspicious contributors contributing to the code base.

2.3 OWASP Top 10 CI/CD Risks[5]

(1) Insufficient Flow Control Mechanisms

Definition: The attacker successfully gained unauthorised access to a critical system within the CI/CD process. However, it became apparent that the compromised system lacked adequate enforcement mechanisms to rigorously approve and review the code or artifacts provided by contributors, leaving a significant gap in the security infrastructure.

Impact:

- The attackers can sneakily add malicious code to a repository branch. This code can either be automatically deployed to the production system or activated manually by the attackers themselves.
- Upload an artifact to the artifact repository, such as a package or container, in the guise of a legitimate artifact created by the build environment and picked up by a deployment pipeline and deployed to production.

Remediation:

- Configure strict branch protection rules
- Limit the usage of auto-merge rules
- Prevent drifts and inconsistencies between the running code in production and its CI/CD origin.

(2) Inadequate identity and Access Management

Definition: This risks stem from the difficulties in managing the vast amount of identities. The identities are identified through personal access token, e-mail, password and so on.

- Overly permissive identities
- Stale identities - Some identities that are not active or no longer require access but have not had their account deactivated.
- External identities - (1) Employees registered with email from a domain not managed by the organization (2) External collaborators are outside the organization's control.

Impact:

- Overly permissive accounts leads to a state where the attacker can compromise any user account on any system within the CI/CD pipeline.

Remediation:

- Consistently checked and connected the accounts of individuals with their access rights, and eliminated any unnecessary access rights that weren't needed for their current tasks.
- Ensure the identities are aligned to the principle of the least privilege, and pre-defined an expiry date for the identities' permissions.
- Prevent the employees from using personal email addresses.
- Avoid the shared accounts. Created the dedicated accounts for each specific context.

(3) Dependency Chain Abuse

Definition: Dependency chain abuse refer to an attacker's ability to abuse flaws relating to how engineering workstations and build environments fetch code dependencies. The build system downloads the malicious package instead of the one intended to be pull. There are four scenarios where the developers might be tricked.

- Dependency confusion - Publication of malicious packages in public repositories with the same names as those private one.
- Obtain the control of the account of the package maintainer in order to upload the malicious version.
- Typosquatting - Publication of similar names to those popular packages.
- Brandjacking - The malicious packages were consistent with the naming convention with the trusted brand.

Impact: Once the malicious code is running, it can be leveraged for credentials theft and move horizontally through a system and network.

Remediation:

- Ensure the packages are not directly pulled through the internet, but through an internal proxy. And disallow pulling directly from external repositories.
- Verify checksum and signature of the pulled packages.
- Lock the packages' version instead of pulling the latest version.
- Installation scripts should not access to sensitive resources in the build process.
- Always ensure projects contain configuration files of package managers.
- The most important is deployment of a quick detection, monitoring and mitigation to avoid further compromise.

(4) Poisoned Pipeline Execution (PPE)

Definition: The attacker access to the source control systems, but without access to the build environment, is able to manipulate the build process by injecting malicious code into the build configuration file. There are three type of PPE, direct PPE (D-PPE), indirect PPE (I-PPE) and public-PPE (3PE).

In the D-PPE scenario, the attackers modify the CI config files either by submitting a PR or directly pushing to the unprotected remote branch. Since the CI pipeline execution is triggered by push or PR events, and the CI execution is defined by CI Configuration file, the malicious commands run in the build node.

In the I-PPE scenario, the pipeline is configured to pull the CI configuration file from a protected branch or CI build is defined by the CI system instead of the in the file stored in the source code. In those cases, the attackers can still inject malicious code into the files referenced by the pipeline configuration file.

In most cases, the permissions of the access to the repository are given to the organization members. However, in the 3PE scenario, the public repositories are allowed the anonymous to contribute. If the CI pipeline runs unreviewed code, the repository is susceptible to the 3PE.

Impact:

- Access to the secret available to the CI job.

- Able to ship code and artifacts further down the pipeline, in the guise of legitimate code build by the build process.

Remediation:

- Ensure that pipelines running unreviewed code are executed on isolated nodes to prevent exposure of sensitive information.
- To prevent the manipulation of the CI configuration file.
- Remove permissions from the users that do not need them.

(5) **Insufficient PBAC (Pipeline-Based Access Controls)**

Definition: Adversary is able to execute malicious code within the context of the pipeline. The pipeline execution nodes have access to the resources or systems within and outside the execution environment.

Impact: Malicious code is able to run in the context of the pipeline. Probability, this attack would lead to the exposure of the secret and confidential data.

Remediation:

- Do not use shared node for pipeline with different levels of sensitivity.
- Where applicable, run pipeline jobs on a separate, dedicated node.

(6) **Insufficient Credential Hygiene**

Definition: CI/CD environments are built of multiple systems communicating and authenticating against each other through verifying the credentials. Insufficient credential Hygiene generally means the overly permissive or the credentials are accidentally existed in CI/CD pipelines and the code repositories. Some cases are due to the unrotated credentials issue.

Impact: The adversary obtains the credentials to deploy the malicious code and artifacts.

Remediation:

- Conform to the least privilege rule and granted the exact set of permission.
- Avoid sharing the same sets of credentials across multiple contexts.
- Using temporary credentials. If using static credentials, better periodically rotate all the static credentials and detect stale credentials.
- Scoping the credentials to specific source IP to ensure the credentials cannot be used outside the environment even if it is compromised.
- Detect secrets pushed to the code repositories.

(7) **Insecure System Configuration**

Definition: Flaws in the security settings and configuration, which often results in easily compromised by The attackers. For example, the overly permissive network access controls allow the attackers to interact with different domain.

Impact: The flaws may be abused by the attacker to manipulate CI/CD flows, and obtain the sensitive tokens.

Remediation:

- Ensure the network access the systems is aligned with the principle of the least access.
- Periodically review all system configuration.
- Ensure permission to pipeline execution nodes follow the least privilege principle.

(8) **Ungoverned Usage of 3rd Party Services**

Definition: 3rd party services are granted access to the organization's assets, such as the CI/CD systems.

Impact: Lack of governance and visibility around 3rd part might allow the write permission on the repository. Then, the flaw is leveraged by the adversary to push the code to the repository.

Remediation:

- Define the scoped context that the 3rd parties are able to access, and with strict ingress and egress filter.
- Established vetting procedures to verify the trustworthiness of the 3rd parties. Prior to being granted access to the environment, the approval of being granted to resources should be verified.

(9) **Improper Artifact Integrity Validation**

Definition: This flaw allows an attacker with access to the CI/CD process to push malicious code or artifacts down the pipeline.

Impact: Execution of malicious code.

Remediation:

- Validate the integrity of resources all the way from development to production.
- Code signing to prevent unsigned commits from flowing down the pipeline.
- Prior to fetching and using 3rd parties, the hash of the 3rd parties should be calculated and cross-referenced against the official published hash of the resource provider.

(10) **Insufficient logging and visibility**

Definition: The risk allow the adversary to carry out malicious activities without being detected. For example, the permission modification and execution of builds.

Impact: Fail to detect a breach may face difficulties in mitigation. Time and data are the most valuable commodities to an organization under the attack.

Remediation:

- First, be familiar with different systems involved in the potential threats.
- Make sure all relevant logs are enables.
- Shipping logs to a centralized location and creating the alerts to detect malicious activities.

2.4 **U.S. Department of Defense Recommend [12]**

1. **Zero Trust Approach**

No user, endpoint device or process is fully trusted.

2. **Strong Cryptographic Algorithm**

Avoid using outdated cryptographic algorithm which poses a threat to CI/CD pipelines. The threat includes sensitive data exposure and keys generated across the CI/CD pipeline.

3. **Minimize the Use of Long-Term Credentials**

4. **Add Signature to CI/CD Configuration and Certify It**

Ensure the code change is continuously signed, and the signature is verified throughout CI/CD process. If the signing identity itself is compromised, it undermines trust.

5. **Two-Person Rules for all Code Updates**

The developer checks in the code which should be reviewed and approved by another developer.

2.5 **in-toto Framework**

in-toto is an end-to-end security framework aim to protect the software supply chain. It provides a variety of mechanism for each steps in the CI/CD to check the integrity of the previous steps. Cryptography and hash function are implemented in in-toto.

Layout is a file format to document the actors, timestamp and actions with each step of the supply chain. It provides the information for downstream steps to verify. This information is the crypto hash of action, actors and some other security related data. Therefore, the downstream steps can verify the integrity based on the hash. This is the high level structure compared to Link Metadata.

Link MetaData is a lower level structure compare to Layout, which provides much detailed information about each action.

2.6 Conclusion

Despite the previously introduced methods seems to address all the security issues existed in the code base and within the CI/CD, some of them may overemphasize one particular approach to address software supply chain security without considering compounding factors that impact risk. Some projects aim at providing single solution that conflates multiple objectives [11]. For example, SAP's Risk Explorer is intended to foster a community of users and contributors to the taxonomy. Users can explore the taxonomy by collapsing and expanding nodes in an attack tree. Detailed information about attack vectors, references, and safeguards is displayed below the tree. This tool covers all potential risk with an entire CI/CD pipeline [9].

Also, the in-toto framework is dedicated to ensure the integrity of the supply chain pipeline and conflates multiple objectives.

The framework implemented in this research is based on SLSA. In the next section, Supply Chain Level Security Artifacts (SLSA) will be introduced and discussed.

3 BACKGROUND

A security framework provides checklist to ensure the integrity of the supply chain. Artifacts that fulfill SLSA requirements endorsed a traceable source of the software provided by trusted providers. SLSA provides trustworthiness of the artifacts to the developers, downstream users [15].

3.1 What is Software Supply Chain?

The software supply chain encompasses a complex network of multiple components, including both first-party and third-party libraries, as well as various processes integral to the development, build, testing, and publication of a software artifact. It serves as the backbone for delivering software products to end-users [12].

3.2 Provenance

SLSA provenance clearly provides the transparent information about the artifacts or the packages. Information such as, who builds this artifact and how the artifact was built from the source. The information will be verified by the package registry or even the customers. The provenance is an attestation in SLSA.

3.3 SLSA Provenance versus SBOM (Software Bill of Materials)

Provenance and SBOM are somehow similar, so they are easily confused. Provenance is used to assess the trustworthiness and security of the processes used to build and deliver the software artifacts 1. By contrast, SBOM focus on listing software components and their versions 2.

Most software is distributed with pre-compiled package, so inspecting the source code is somehow impossible. SBOM provides a mapping between binary code and the materials of a software. For example, the version of the software will be documented in SBOM. However, the developers cannot trust the pre-compiled artifact is actually derived from the authentic source code. Usually, the current trend will tend to build the reproducible artifact which means the binary code is able to be rebuilt from the source code if using the given hardware. Then, the developers can verify the integration through comparing the hash between pre-compiled package and rebuilt artifact [4].

Listing 1: SLSA Provenance

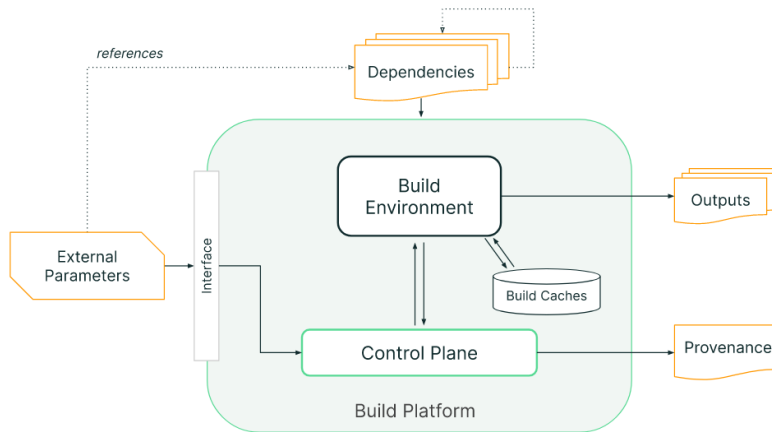
```
[Software Build Provenance]
Build Date: 2023-09-01
Build Environment: Secure, Isolated Environment
Signing Authority: Trusted Certificate Authority (CA)
Signature Verification: Passed
[Supply Chain Processes]
Code Review: Multi-stage code review by security experts
Dependency Scanning: Automated scanning for known vulnerabilities
Build Automation: Continuous Integration/Continuous Deployment (CI/CD) pipeline
Deployment: Automated deployment to secure servers
[Organizations Involved]
```

Development Team: Responsible for code development
 Security Team: Responsible for security reviews and scanning
 Operations Team: Responsible for deployment

Listing 2: SBOM

```
[MyApp (v1.0)]
Frontend Framework (v2.3)
Database Connector (v1.1)
Authentication Library (v3.0)
Logging Utility (v1.2)
```

3.4 Build Model



1. The tenant (developers) provide specific external parameters, like the version of the application and the reference to the dependency.
2. The control plane receives the external parameters, then fetch the necessary build scripts, configuration, and dependencies based on the external parameters.
3. The control system sets up the isolated environment for the build.
4. Finally, the model outputs the artifacts. If the build platform follows SLSA build level 2+, then the provenance will also be generated from the control systems.

3.5 SLSA Security Level

Currently, SLSA have defined 4 levels (LV.0 - LV.3), which will be briefly described in Table 1, and working on level 4.

3.6 Policy Engines

Policy is a logic code, which will decide whether the action can be operated. There could be nested policies within on policy. For example, the consumer want to check if they are allowed to access the cloud and download the data from the cloud. They will if their provided password valid and whether the source IP is not within the banned IP list. Then, the policy to check whether the password is valid is based on the correct password format and the password is stored in database.

Policy Engine is able to adopt the results generated from other tools, like the checker in the framework of our research or the malicious code injection scanner. The Trusted Wrappers provide and interface to convert the results from the original format into the logic. If the policy is verified, it will become a fact and store in the Transparency log. Everyone can directly monitor the facts store within the transparency log, which prevent the logic from being compromised. The transparency log is untrusted to the consumers, which means the consumers will not trust the maintainers of the log [4].

Our framework implements the policy engines combined with the result from the check in order to provide a clear view and easy way to understand the security requirement of the consumers.

Table 1: SLSA Security Levels

Level	Requirements	Focus
Build L0	None	No security practices are in place.
Build L1	With provenance	Basic security practices are followed, such as code review and basic dependency scanning.
Build L2	Signed provenance, generated by a hosted build platform	More comprehensive security practices are implemented, including in-depth code review, vulnerability scanning, and build verification.
Build L3	Hardened build platform	The highest level of security is maintained, with strict adherence to security practices, automated testing, and supply chain integrity checks.

4 METHODOLOGY

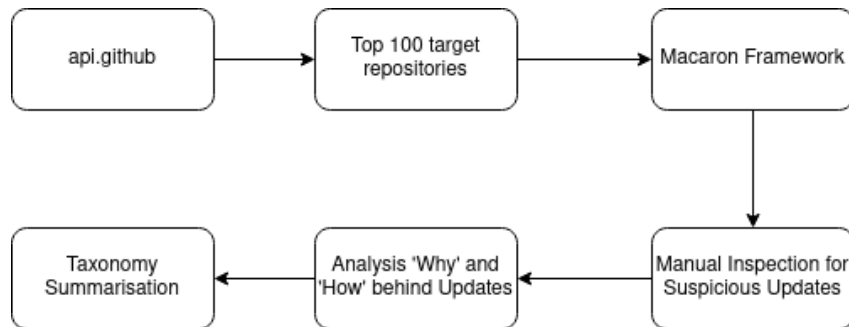
According to previous works from other researchers, most of the works mainly focus on detecting the malicious code within the source code base. Despite this traditional method might be able to discover the suspicious code through scanning, it is an inefficient process if the code base is frequently updated. Also, the method can only promise the malicious code within the whitelist to be detected. However, the attackers usually figure out possible payloads or injection methods to bypass the automatic scanning, like the Trojan source attacks [2] we introduced before. Some works recommend to use machine learning method to detect malicious code as an outlier [6]. How if the attacker try to bypass the training model with obfuscation method? Finding out proper features for training a model is time-consuming. In Paper [19], the author discover a method to modify the forward function in deep learning model, which will indirectly poison the downstream developers who build their model based on the incorrect pretrained weight. Therefore, we are thinking of whether developing a method to grapple with malicious code injection is possible. If our framework can make sure the contributors are not being compromised and the CI/CD configuration follows the safe requirements suggested by security experts, the developers can trust the dependencies that are used in their project. In order to efficiently check the safety of the artifacts, our research will focus on contributing to the Macaron Framework which is based on SLSA requirements.

Three major questions will be addressed in this research:

RQ1: What are the top five frequently emerging risks that exist within popular Python and Java repositories?

RQ2: What is the intent behind these suspicious updates from contributors in open source projects?

RQ3: How can maintainers enhance the security of their CI/CD pipelines based on the findings and recommendations from our research?



4.1 Research Aims

This research is aim to contribute the Macaron framework, then examine top 100 Python and Java Git repositories with this framework. The statistical findings will be further concluded through the further

process. Also, the results will provide the developers and maintainers with a good understanding of the vulnerabilities existed in their CI/CD pipelines configuration. Furthermore, the reason behind these unsafe update will be deeply investigated.

4.2 Research Objectives

1. First step: Fetch the repositories build from Python and Java with the top 100 most stars.
2. Second step: Input the repositories name into the Macaron Framework to analysis.
3. Third step: Summarize the outputs generated by Macaron and visualize the results with graphs.
4. Fourth step: Manually inspecting the code base from the potentially problematic repositories due to not comply with the requirements from the SLSA.
5. Fifth step: Document and investigate the reason for this suspicious updates.

5 TIMETABLE AND PLAN

We split the research plan into three phases, also, providing a timetable for understanding the research schedule.

5.1 Phase One

Defining Unsafe Updates: Building on similar research conducted in the JavaScript ecosystem (<https://ieeexplore.ieee.org/document/8805698>), the project's first phase involves defining what an unsafe update means within the context of Python and/or Java. Typically, an unsafe update could be one that introduces breaking changes, negatively affects performance, opens up security vulnerabilities, or adds incompatible API changes.

In order to discover different type of unsafe updates and the target victim, we will review related work, articles, and CI/CD attack events to support the understanding of the CI/CD attacks existed nowadays.

5.2 Phase Two

Implementation of Safety Checks: Next, we will extend the Macaron framework's functionality by implementing additional safety checks for these unsafe updates. Macaron is an extensible checker framework for supply chain security and CI/CD services, such as GitHub Actions. It allows adding new checks as Python modules and provides intermediate representations specifically designed for CI/CD services to facilitate verifying new properties. We will begin from implementing SLSA Level 4 check including Two Person Review, Verified History, and Retained indefinitely.

A malicious and suspicious event on GitHub is conducted by one user account "pastramahodu" which forked nearly 2000 repositories. The researcher find out the attacker try to use script to automatically fetch and detect the target file type, such as package.json, sh, and so on. Then, the script creating pull request against each victim repositories to compromise the repositories that are not configured properly. The malicious code injection will be triggered in CI pipelines to exfiltrate repository name and hostname back to webhook server [7]. Therefore, we decide to implement the following three checks to avoid the malicious code being merged into the branch.

The Two Person Review check ensures each pull request is reviewed by at least one authentic reviewer. This check deal with the situation that someone wants to bypass the review and directly merge their code into the branch. Our method will fetch all the pull requests from the branch specified by the users of Macaron Framework.

The Verified History ensures at least one strong authenticated participant of the revision's history. The identities should be authenticated through two-step verification. The first step usually verified the password, and second step might be verified through SMS or Email. This check grapple with the situation when the identities' account are being compromised.

Retained indefinitely will check if the commits are preserved for 18 months; therefore, the consumers can trust the artifacts they are going to use in their applications are not being modified by suspicious contributors.

The remaining session, we will keep contributing to the framework, but only implementing a specific version on the phase three.

5.3 Phase Three

Empirical Analysis of Real-World Projects: With the safety checks in place, the final phase of the project is an empirical study conducted on GitHub to ascertain the frequency of unsafe updates occurring in Python and/or Java projects. By understanding the 'how' and 'why' behind these updates, developers can adopt more informed, proactive strategies in their coding practices. We will design a script to fetch our data, which will be designed to perform the Python script Listing 3.

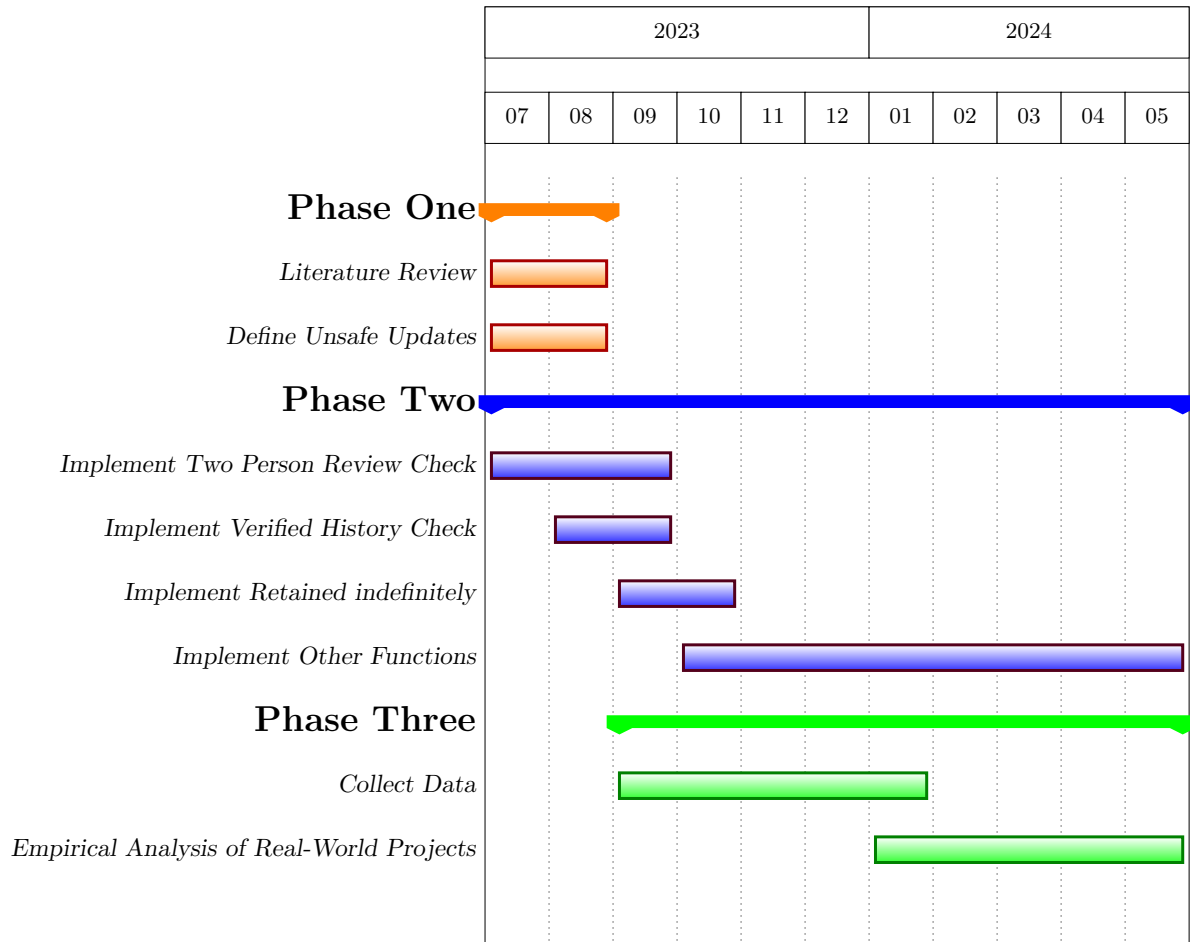
Based on the suspicious commits detected by the Macaron Framework, we will further check the code base through static checking tools, then dismissing the false positive case through manual inspection. Considering the Python malicious package, typo-squatting is the most frequent case to compromise the consumers who do not carefully check the packages they download. Usually, the attackers' main goal is using the injected code to connect to their webhook server or build up a reverse shell, which will cause sensitive data exfiltration. And the malicious code are prone to be injected in the setup.py which is triggered when running the installation script. Therefore, our static analysis method can include this rule [1].

Furthermore, we will build graph and even attack tree to classify our finding and visualize the result. In this way, the developers can easily understand what they should focus and improve in their current configuration of the CI/CD pipeline or the future projects.

Listing 3: Fetch Top 100 Java and Python Repositories

```
1  import requests
2  import os
3  import sys
4  access_token = os.getenv("GITHUB_TOKEN")
5  if access_token is None:
6      print(f"Warning: {access_token} environment variable is not set.")
7      sys.exit(1)
8  url = "https://api.github.com/search/repositories"
9  language = ['java', 'python']
10  params = {
11      'q': f'language:{language}', # Filter for repositories with more than 0
          stars
12      'sort': 'stars', # Sort by stars
13      'order': 'desc', # Sort in descending order (most stars first)
14      'per_page': 100, # Number of results per page (max is 100)
15      'page': 1, # Page number (start with 1)
16  }
17  headers = {
18      'Authorization': f'token {access_token}',
19  }
20  # Send the GET request to the GitHub API
21  response = requests.get(url, params=params, headers=headers)
22
23  # Check if the request was successful
24  if response.status_code == 200:
25      # Parse the JSON response
26      data = response.json()
27
28      # Print information about the top repositories
29      for idx, repo in enumerate(data['items'], start=1):
30          print(f"{idx}. {repo['name']} - {repo['html_url']} - {repo['stargazers_count']} stars")
31  else:
32      print(f"Error: {response.status_code} - {response.text}")
```

5.4 Timeline



6 CONCLUSION AND EXPECTED OUTCOMES

6.1 Conclusion

In this research proposal, we analyse different methods developed in the previously related works. And we investigate the pros and cons of these methods in order to figure out the research gap that we can improve in our research. Also, we provide the overview for three main attack surfaces in CI/CD pipeline, then discussing some potential vulnerabilities and counter measure on the security of the software supply chain. And we summarise the Top 10 risks in order to provide a clear view of which risks should be focused more and tackle with first. Since our research tool is developed based on the SLSA, we introduce key components and concepts of this framework as a prerequisite. In the final section, the method and the plan of our research are being discussed. Our method tries to improve the method to detect suspicious updates that is not achieved in previous research.

6.2 Expected Outcomes

The expected outcomes of this research will be contributing some checks properly to the Macaron Framework. Also, finding out vulnerabilities within the popular repositories. Our work expects to find out if the CI/CD pipeline configuration or the protection mechanism is strong enough to avoid unexpected attacks. Taxonomy technique mentioned will be implemented in our research to classify the discovered vulnerabilities, and finding out the potential main goal of the attack according to these vulnerabilities [13]. Finally, our research will expect to discover the reason behind these malicious updates.

References

- [1] Bertus. Detecting cyber attacks in the python package index (pypi), October 13 2018.

- [2] Nicholas Boucher and Ross Anderson. Trojan source: Invisible vulnerabilities, 2023.
- [3] ESLint. Postmortem for malicious packages published on july 12th, 2018.
- [4] Andrew Ferraiuolo, Razieh Behjati, Tiziano Santoro, and Ben Laurie. Policy transparency: Authorization logic meets general transparency to prove software supply chain integrity. In *Proceedings of the 2022 ACM Workshop on Software Supply Chain Offensive Research and Ecosystem Defenses*, pages 3–13, 2022.
- [5] OWASP Foundation. Owasp top 10 ci/cd security risks, 2023.
- [6] Kalil Garrett, Gabriel Ferreira, Limin Jia, Joshua Sunshine, and Christian Kästner. Detecting suspicious package updates. In *2019 IEEE/ACM 41st International Conference on Software Engineering: New Ideas and Emerging Results (ICSE-NIER)*, pages 13–16. IEEE, 2019.
- [7] Yehuda Gelb. Mass scanning of popular github repos for ci misconfiguration, May 2023. Published in *checkmarx-security*.
- [8] Piergiorgio Ladisa, Henrik Plate, Matias Martinez, and Olivier Barais. Sok: Taxonomy of attacks on open-source software supply chains. In *2023 IEEE Symposium on Security and Privacy (SP)*, pages 1509–1526. IEEE, 2023.
- [9] Piergiorgio Ladisa, Serena Elisa Ponta, Antonino Sabetta, Matias Martinez, and Olivier Barais. Journey to the center of software supply chain attacks, 2023.
- [10] Luis Mastrangelo, Luca Ponzanelli, Andrea Mocci, Michele Lanza, Matthias Hauswirth, and Nathaniel Nystrom. Use at your own risk: The java unsafe api in the wild. *ACM Sigplan Notices*, 50(10):695–710, 2015.
- [11] Marcela S. Melara and Mic Bowman. What is software supply chain security?, 2022.
- [12] U.S. Department of Defense. Defending continuous integration/continuous delivery (ci/cd) environments. PDF document, 2023.
- [13] Marc Ohm, Henrik Plate, Arnold Sykosch, and Michael Meier. Backstabber’s knife collection: A review of open source software supply chain attacks. In *Detection of Intrusions and Malware, and Vulnerability Assessment: 17th International Conference, DIMVA 2020, Lisbon, Portugal, June 24–26, 2020, Proceedings 17*, pages 23–43. Springer, 2020.
- [14] Sean Peisert, Bruce Schneier, Hamed Okhravi, Fabio Massacci, Terry Benzel, Carl Landwehr, Mohammad Mannan, Jelena Mirkovic, Atul Prakash, and James Bret Michael. Perspectives on the solarwinds incident. *IEEE Security & Privacy*, 19(2):7–13, 2021.
- [15] SLSA Collaboration. Supply chain levels of software assurance (slsa). Accessed on September 6, 2023.
- [16] Duc-Ly Vu, Fabio Massacci, Ivan Pashchenko, Henrik Plate, and Antonino Sabetta. Lastpymile: identifying the discrepancy between sources and packages. In *Proceedings of the 29th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, pages 780–792, 2021.
- [17] Ying Wang, Bihuan Chen, Kaifeng Huang, Bowen Shi, Congying Xu, Xin Peng, Yijian Wu, and Yang Liu. An empirical study of usages, updates and risks of third-party libraries in java projects. In *2020 IEEE International Conference on Software Maintenance and Evolution (ICSME)*, pages 35–45. IEEE, 2020.
- [18] Gentoo Wiki. Gentoo wiki incident report, June 2018.
- [19] Tianhang Zheng, Hao Lan, and Baochun Li. Be careful with pypi packages: You may unconsciously spread backdoor model weights. *Proceedings of Machine Learning and Systems*, 5, 2023.