Supply Chain Vulnerability Analysis

Varshini M, *Fellow, IEEE*, Srishti Mathur, *Fellow, IEEE*, Nagasundari S, *Member, IEEE*,

[[1]](#footnote-1)

***Abstract*— Supply chain attacks have become a notable menace to the software sector, resulting in extensive harm and penetration. These attacks capitalize on vulnerabilities within the security protocols of third-party suppliers, resulting in widespread security breaches that can have extensive repercussions for both the suppliers and cloud service providers associated to them. A considerable amount of investigation has been carried out regarding the origins of these attacks, yielding detailed accounts of significant incidents such as the SolarWinds breach. Nevertheless, there is still a necessity for comprehensive visualization tools to aid in predicting future vulnerabilities based on past patterns. To address this gap, this research introduces a distinctive approach to monitoring the evolution of supply chain attacks targeting major third-party suppliers. We also incorporate dependency-oriented vulnerability assessments to enhance threat identification and mitigation.**

***Index Terms*— *supply chain attacks, vulnerabilities, third-party suppliers, visualization, npm audit, mitigation***

# I. INTRODUCTION

The software supply chain (SSC) represents a sophisticated and intricate web of entities, including companies, individuals, and organizations, engaged in the creation and dissemination of software products, with a primary goal of leveraging existing software components. The SSC assumes a crucial position in the realm of software application development, distribution, and utilization on a global scale, a significance that has been magnified by the escalating complexity characterizing software development endeavors in recent times [1].

A software vulnerability is an imperfection within a software system that can be used by an unauthorized agent to take over a system. This allows an attacker to breach sensitive information, create a botnet with the system, install a backdoor, or carry out various types of malware attacks[2].

## A. Supply chain attacks: A Growing Threat

One notable and relatively recent software vulnerability pertains to the emergence of supply chain attacks, a trend that has garnered noticeable attention within the software industry. The open-source community offers numerous modules and

packages that are extensively embraced by businesses globally, including those integrated into a supply chain.

Nevertheless, the drawback of open-source lies in its inherent vulnerability. This is largely due to the fact that it is authored by individuals who might lack the necessary expertise or financial resources to ensure its complete security.

This loophole in one’s security framework arises because the open-source packages the company incorporates could have undisclosed dependencies. This was precisely the case with NotPetya: a variant of conventional malware, NotPetya managed to breach systems worldwide by exploiting a widely utilized open- source accounting software. Consequently, it spread rapidly, causing turmoil in Ukraine and various prominent nations like the U.K., France, Germany, Russia, and the U.S [3], [4].

The study, entitled "The State of Supply Chain Defense 2023 Annual Global Insights Report" and released by BlueVoyant, indicated that the average number of supply chain failures grew from 3.29 in 2022 to 4.16 in 2023, with a projected 26% increase in supply chain breaches impacting organizations between 2022 and 2023 [5]. A study predicts that by 2025, 45% of organizations will most likely have been victim to a software supply chain attack [6].

Organizations persist in their inadequate preparedness to fully grasp the extent and characteristics of threats emanating from third-party providers. A notable obstacle lies in the assurance of timely mitigation of threats by suppliers in the supply chain, along with the effective dissemination of information pertaining to vulnerabilities and security concerns [7]. In a supply chain assault, the assets of the supplier that are the target of the assault may encompass software, hardware, data, processes, and individuals. The majority of the 24 supply chain assaults (spanning from January 2020 to early July 2021) scrutinized in the 2021 ENISA Threat Landscape for Supply Chain Attacks report are supply chain assaults relating to software. More specifically, 62% of these 24 supply chain assaults have employed malicious software; 50% have been linked to well-known APT groups (such as APT 29), and 58% have been directed towards the theft of clients' data [8].

The well-documented 2013 breach targeting the prominent US retail entity Target serves as a prominent illustration of lateral movement [9]. SolarWinds, a SaaS provider, is the developer of Orion, software designed to manage and monitor business infrastructure. In 2020, cybercriminals buried a backdoor within a plugin distributed as an Orion update. Approximately 18,000 customers, including technology firms and government agencies, had their security compromised by the malicious update [10].

## B. Elements and Methods of Supply Chain Attacks

Elements of a supply chain include third-party software used to interface with enterprise software, open-source platforms, cloud services, and past and present suppliers and vendors that have access to a company’s data and systems [11]. Although such attacks can take various forms, the two most common ones with respect to a detrimental effect on the cloud infrastructure and services are:

## **Harmful Code or Hardware Insertion:** Attackers insert harmful code or hardware into the production or distribution process[11].

## **Compromise with Third-Party Service Providers:** Cloud services frequently rely on third-party vendors for functionality such as authentication, encryption, and content delivery. If one of these vendors is compromised, it can have an impact on the overall security of the cloud service [11].

## C. The Cloud and Its Associated Risks

The cloud leverages software capabilities for agility, setting it apart from traditional data center resource provisioning. The components of cloud services, particularly Software-as-a- Service (SaaS) applications, are composed of loosely coupled services. These software programs are often developed by a large, dispersed group of developers worldwide. Cloud Service Providers (CSPs) typically do not build software from the ground up but rather assemble cloud applications using components from global vendors [12].

Recent studies have indicated that at least 80% of a typical SaaS application consists of assembled parts, each presenting varying levels of risk. In 2015, for instance, 1.8 billion vulnerable open-source components were downloaded, with at least 26% of the most common open-source components containing high-risk vulnerabilities[12].

The Wiz Research team conducted an extensive analysis of permissions granted to third-party vendors in cloud environments, focusing on direct exposure via IAM roles. Their research examined the permissions provided to over 1,300 AWS accounts across more than 40 popular third-party vendors. The key findings reveal that 82% of companies provide highly privileged roles to third-party vendors, and 76% have roles that allow for full account takeover. Additionally, over 90% of cloud security teams were unaware of the high permissions granted to these vendors. The study also found that the majority of these permissions are unnecessary, as vendors do not actually need them, and customer teams are often unaware they have granted such access[13].

According to IBM's 2022 Cost of a Data Breach Report, supply chain compromises accounted for 19% of breaches. The average overall cost of a third-party breach was $4.46 million, or 2.5% more than the average cost of other types of breaches. Furthermore, detecting and mitigating third-party breaches took an average of 26 days longer than the global average for all types of breaches [14].

#### D. Motivation and Contribution

While extensive research has been conducted on the detection and prevention of supply chain attacks, this paper primarily focuses on providing detailed visualizations of a software vulnerability dataset. This dataset encompasses a large number of vulnerabilities spread across various common third-party vendors over the years 2014-2022, offering both historical context and potential future trends. By analyzing this dataset, the aim is to reveal patterns and insights that can help in understanding the evolution of software vulnerabilities over time.

In addition to this analysis, the paper explores the results and usage of the common vulnerability detection tool, npm audit. Although npm audit is already a powerful tool for identifying vulnerabilities in Node.js packages, this paper proposes enhancing its functionality by adding a new feature—a vulnerability details visualizer. This visualizer will present detailed information about identified vulnerabilities in a more accessible and interpretable format. By doing so, it will simplify the analysis process, enabling third-party vendors to better understand the specifics of each vulnerability and to implement appropriate measures to counteract and resolve these issues effectively.

Moreover, the study encompasses the creation of a predictive model for evaluating the seriousness of identified vulnerabilities, leveraging historical data and machine learning methodologies to anticipate the potential consequences of new vulnerabilities. In addition to assessing severity, the model will offer recommendations for detection and mitigation strategies, aiming to raise awareness among vendors and assist them in adopting proactive measures to mitigate the adverse impacts of supply chain attacks in computing ecosystems.

By integrating intricate visual representations with predictive analyses, this strategy aims to enrich the comprehensive comprehension and handling of software vulnerabilities empowering third-party vendors to enforce more efficient security protocols ,ultimately diminishing the likelihood and consequences of supply chain attacks.

The rest of this paper is structured as follows.

Section II introduces related work. Section III describes

the vulnerability dashboard. An added functionality in the npm audit tool, is proposed in Section IV. Section V proposes a ML model for prediction of attack severity. Section VI summarizes the results and concludes the paper along with future scope.

# II. Related Works

## A. Risk assessment strategies and current limitations in Security Control

The recent evolution of cyber assaults against the software supply chain has revealed that cyber actors are increasingly focusing their attacks on open-source projects. An important point that should be factored in more detail when assessing the impact of those vulnerabilities is their exploitability and the degree of damage each vulnerability might cause—one which could be aided by relying on the resources provided by others, e.g., the U.S. National Vulnerability Database (NVD) and Software Composition Analysis (SCA). When it comes to resources in the form of publicly available knowledge repositories, however, NVD, GitHub Security Advisories, and the Sonatype OSS index are quite useful in aiding the known vulnerabilities process. This study pushes for future research efforts to focus on the development of more comprehensive frameworks of risk assessment and the smooth integration of robust vulnerability management practices across the entire software supply chain [15].

The weaknesses within the software supply chain by more specifically examining the SolarWinds breach and, simultaneously, providing plausible solutions. The important steps taken include keeping the software patched up and disabling USB ports that will mitigate vulnerabilities such as use-after-free and privilege escalation in the identified solutions of CVE-2020. Future approaches underline the necessity of living up to SOC2 standards—not only managing vendor risk but also coming from internal governance—reinforced with penetration testing practices, all well-customized for the industry [16].

# III. Vulnerability Visualization

In 2022, the number of Known Exploited Vulnerabilities more than doubled by the end of the year, from [311 CVEs to](https://vulncheck.com/blog/2022-cisa-kev-review) [868](https://vulncheck.com/blog/2022-cisa-kev-review) [17]. This research began with a careful data collection process. Researchers drew on an extensive data set from the CISA Known Exploited Vulnerabilities catalog for 2022[18]. This catalog is a compilation of documented security vulnerabilities that have been successfully exploited, as well as vulnerabilities associated with ransomware campaigns. This dataset includes detailed information of the security vulnerabilities from 125 unique vendors across a period of 8 years from 2014 to 2022.

## A. Dataset Explanation

This dataset comprises 12 significant key attributes.

**CVE Identifier:** An alphanumeric string that designates a Publicly Disclosed vulnerability, following the format outlined in the CVE Record Format established by CNAs [19].

**Vendor Project:** A third-party entity linked to the vulnerability, typically an external individual or organization offering a service or technology under a contract.

**Product:** Products provided by the vendor which are related to the vulnerability, they allow understanding security weak points of various products.

**Vulnerability Name:** This is a descriptive title for the vulnerability where it is identified with its type and characteristics. This helps in understanding the scope of the vulnerability and its possible consequences.

**Short description**: Briefly describe a little about the vulnerability: type, the software affected, and the results of the attack that could happen.

**Publication Date:** The public release date of a vulnerability is important to realize exactly when a risk was open to the world.

**CVSS:** Common Vulnerability Scoring System represents an openly available framework that scores and enumerates attributes of security vulnerabilities in information systems.

**CWE:** CWE is Common Weakness Enumeration for the weakness and vulnerabilities classified in hardware and software literature.

**Attack Vector:** Describes how an attacker could possibly exploit the vulnerability. This will greatly contribute to determining the accessibility and exploitation possibilities and give vital insights on how to develop mitigation strategies.

**Severity Levels:** The severity levels are normally divided in line with the degree of consequences that the vulnerability can cause to the impacted systems and the extended organizational environment. In the given study they are classified into medium, high, and critical.

**Cloud Component:** Specify which cloud component is affected by the vulnerability so that mitigation strategies are put in place for that component only.

**Category of vulnerability:** This involves the class or grouping to which vulnerability belongs.

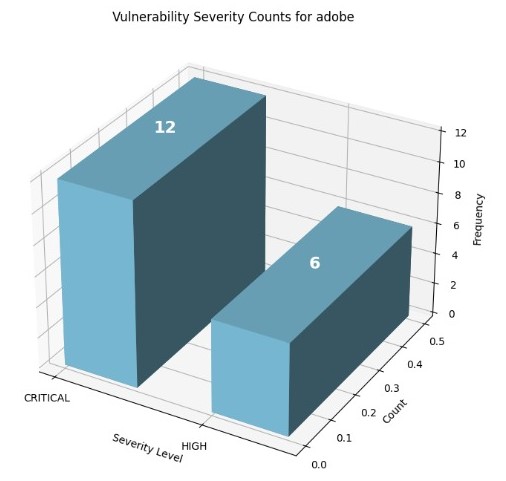
## B. Data Visualization

Data visualization is essential for interpreting and sharing cybersecurity data. Through converting unprocessed data into visual representations like charts, graphs, and maps, data visualization simplifies intricate information, emphasizes trends, and uncovers insights that may go unnoticed.

### 1) Vulnerability Severity Counts

The chart gives a summary of the quantities pertaining to vulnerabilities, based on their severity levels: High, Medium, and Critical. We label vulnerabilities against their respective CVSS scores.

The vulnerabilities with a High Severity rating, thus part of the 7.0 to 8.9 severity spectrum in the CVSS scale, are those considered critical vulnerabilities whose exploitation can result in severe harm. The Medium Severity vulnerabilities have an intermediate score ranging between 4.0 and 6.9, with moderate levels of vulnerabilities that may be dependent on user interaction or other prerequisites to happen. Critical severity vulnerabilities are rated in the 9.0 to 10.0 range and represent very high vulnerabilities: easily exploitable, potentially very harmful if exploited.

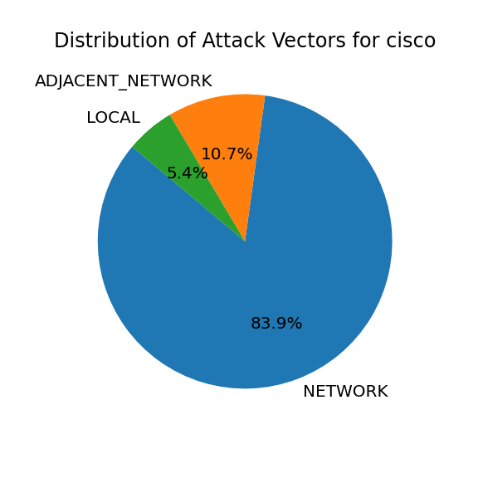


**Fig. 1.** Graph for Vulnerability Severity Counts

An approach treating the vulnerabilities based on their criticality forms the basis for prioritizing these vulnerabilities for remediation. High- and critical-severity vulnerabilities are those which require the quickest possible fixes to cut down on the exposure to possible threats against systems, security, and functionality, and concerned networks linked with them.

## 2) Distribution of Attack Vectors

The chart, in its vulnerability distribution by attack vectors, represents an important dispelling of confusion between Network and Local vulnerabilities.



**Fig. 2.** Pie Chart for Distribution of Attack Vectors

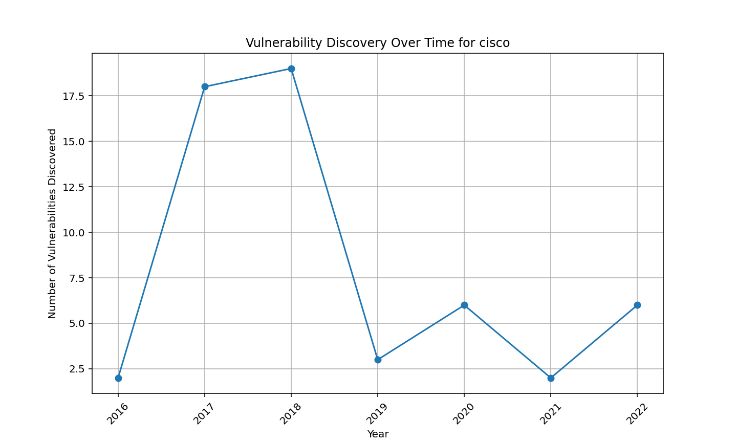
In that network vulnerabilities are remotely exploitable, not requiring physical access to the specific device, the potential for exposure from every point on the globe to possible actors represents a serious threat. This need emerges considering that network vulnerabilities are always present.

In marked contrast, local vulnerabilities involve either physical or logical access to a device and present an entirely different set of issues and concerns. While the risks posed by local vulnerabilities are relatively lower, this is still pretty significant in an environment where insider threats are a

primary concern.

### 3) Vulnerability Discovery Over Time

This line chart will show clearly the progress of the vulnerability discovered over time, thereby helping in visualizing the trends in vulnerability detection within a given period.



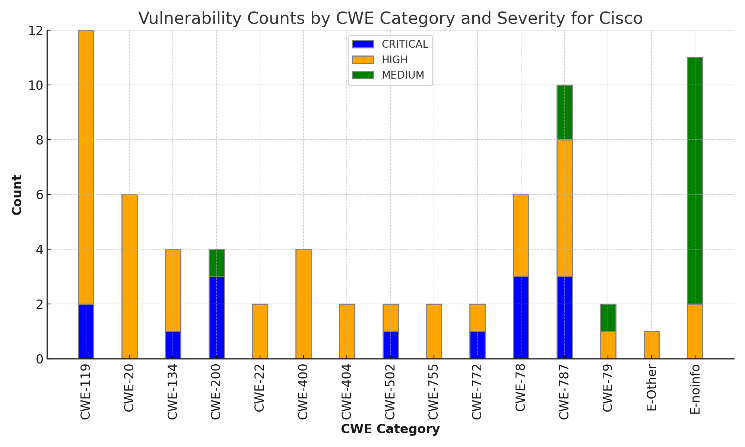
**Fig. 3.** Trend Analysis of Vulnerabilities over Time

The periods of time associated with higher numbers of vulnerability discoveries may have peak years in which the majority are premiums, which can be linked to a number of factors: from amplified research through improved detections to heightened threat activity in the cyber domain. The opposite—an overall decreasing trend of vulnerability discovery—might point to increased security practices, reduced new vulnerabilities, or a shift in the cyber domain strategies of malicious actors.

These trends in the discovery of vulnerabilities may be used by any given organization for the estimation and preemption of upcoming security challenges and, therefore, marshal resources effectively for strategic deployment. For instance, if the rate of discovery for vulnerabilities keeps increasing constantly, an organization may decide to increase its security research and development investments in order to align with the emerging threats and vulnerabilities. Using historical data on vulnerability discoveries, any organization can make very informed decisions on financial planning and resource allocation for security research, development, and mitigation efforts, enhancing its overall cybersecurity posture and resilience against threats and cyber-attacks.

### 4) Vulnerability Counts by CWE and Severity

The bar chart expresses the count of vulnerabilities, differentiated by their CWE categories and colored by severity.

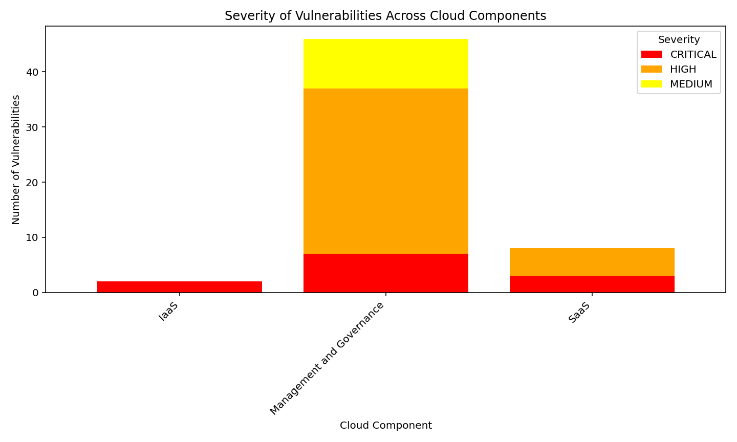


**Fig.4.** Graph for Vulnerability Counts by CWE Category

Knowing the exact CWE categories helps to recognize common developer mistakes and provides a lead for secure development. Color differentiation by severity within each CWE category helps recognize which weaknesses are more critical. Knowing the trending CWE categories and their severity distribution enables organizations to focus on the most critical ones and the root cause of problems.

5) Severity of Vulnerabilities of Cloud Components

A bar chart showing the vulnerability distribution by severity ranking across the different components of cloud systems can be drawn as shown.



**Fig. 5.** Graph for Severity of Vulnerabilities across Cloud Components

The sheer complexity of securing IT environments has risen already because too many organizations have embraced cloud solutions [20]. Prioritization by severity can be accomplished using visualization tools. Effective prioritization would be required as resources would always be at a premium while organizations would be faced with high-risk vulnerabilities needing to be mitigated without much delay [21].

**IaaS (Infrastructure as a Service)**

IaaS sells virtualized computing resources that enable customers, via the Internet, to rent and manage important components of infrastructure without having to own physical hardware. The main thing on offer in this case is virtual machines, storage, and networking resources—users retain control over operating systems, applications, and configurations [22]. Some vulnerabilities include misconfigured VMs, cloud storage problems, no cloud infrastructure visibility, shared tenancy issues, and insider threats from employees at cloud providers.

Mitigation Strategies: The state-of-the-art access controls and multi-factor authentication should be ensured against those vulnerabilities. NSGs should also be working on both incoming and outgoing traffic to VMs. VM disks and data at rest have to be encrypted. Monitoring, logging, and security incident response should be established. Regular security assessments and penetration testing are also highly recommended [23].

**PaaS (Platform as a Service)**

It is the platform that PaaS offers, including not only the infrastructure—virtual machines and storage—but also development and management tools that operate at a high level of abstraction from the underlying infrastructure. Basically, the role of PaaS would be to provide a platform for creating, testing, and running applications. As designed, it is between Infrastructure as a Service and Software as a Service, so PaaS will be perfect for network applications, APIs, and general software solutions [22]. Examples include platform vulnerabilities in the underlying infrastructure, application vulnerabilities in custom code and configurations, insecure interfaces and API, misconfigured system settings, and exposure of sensitive data and secrets [24].

Mitigation Strategies: Secure coding practices, regular application testing, and setting up container image security, combined with the visibility and control of cloud use by a Cloud Access Security Broker, workload security by means of Cloud Workload Protection Platforms, and threat modeling for risk identification and mitigation are recommended for mitigation against these PaaS vulnerabilities [25].

**SaaS (Software as a Service)**

It is a service that provides full software applications over the Internet, accessed by users through web browsers only [30]. The software is maintained and hosted by the service provider. This ranges from a very wide spectrum of applications from empowerment tools down to very specialized business software tools [22]. Some of the vulnerabilities include unauthorized access and account takeover, insider threats from employees at the SaaS provider, data breaches and loss, and compliance and audit failures [24].

Mitigation Strategies: Regarding SaaS vulnerabilities, proper identity and access management controls should be in place; the data at rest and in transit needs to be encrypted; configurations should securely follow vendor-recommended best practices; incident response and disaster recovery plans should be framed in a proper way; and compliance can be assured through regular audits [24].

**Management and Governance**

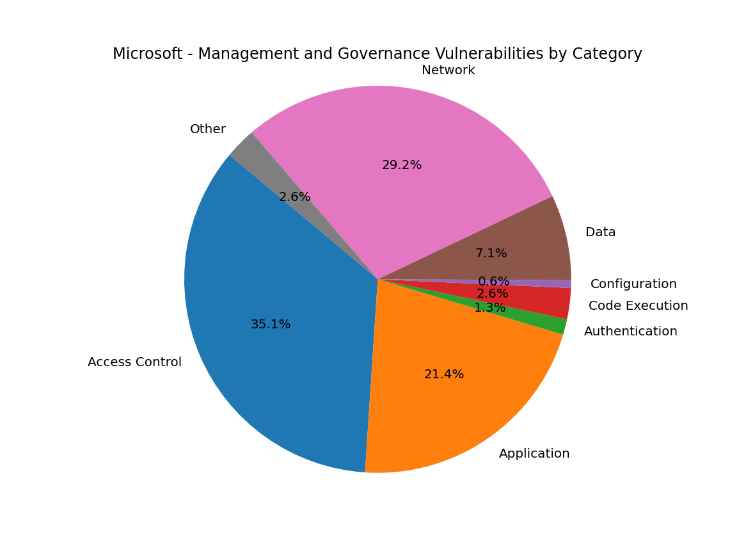
This layer deals with orchestration and management across cloud services, security policies, and compliance. In this case, the vulnerabilities to management and governance may source from the following areas: not transparent across cloud environments, misalignment of business and IT teams, and poor security policies and controls [26].

Mitigation Strategies: The strategies that offset at the management and governance layer for mitigating vulnerabilities are aimed at developing a comprehensive strategy and roadmap on cloud security, baseline setting, security requirements, and implementation on cloud services; implementation of the CSPM; ensuring the identification of risks, clarity in roles and responsibilities, and accountability of cloud security; provisions for cloud security training and awareness among employees [26].

### 6) Vulnerabilities by Category for Each Cloud Component

Technical analysis of the data from the CISA Known Exploited Vulnerabilities catalog allows the trending of clear lines regarding the severity of the vulnerabilities and their distribution across different cloud components. All these trends are important to understand for formulating pinpointed security measures. Control is ensured through strong security measures at the network and data levels in the IaaS environment, by secure configuration and following secure development practices in PaaS, through stringent application security in SaaS and by a comprehensive security approach in management and governance.

There exist 11 key divisions according to the dataset within cloud elements that, while interconnected, jointly influence the entirety of cloud architecture. These divisions encompass Network, Code Execution, Application, Data, Identity and Access Management (IAM), Configuration, Access Control, Cryptography, Authentication, Denial of Service (DoS), and Hardware. The pie charts display the distribution of vulnerability categories for each cloud component. The vulnerability categories across different impacted cloud component



**Fig. 6.** Pie chart for vulnerabilities by category

*Vulnerabilities in IaaS (Infrastructure as a Service)*

Vulnerabilities in IaaS networks are due to non-patched hardware, under-configuration, or no proper monitoring; this actually results in network sniffing, man-in-the-middle attacks, or unauthorized access. Vulnerabilities like code execution can make way for an outbreak in VMs and containers, and these denote an outbreak in the system and data leak. Application vulnerabilities, such as code injection and insecure APIs, further increase the threat exposure level. This can cause unauthorized access and compliance issues through insecure storage and weakness in encryption. IAM vulnerabilities might cause a virtual resource attack, privilege escalation, and resource and identity spoofing. Misconfigurations expose the different areas to threats, whereas weak access control may lead to data theft and service disruption. Cryptographic vulnerabilities they may result in the compromise of information due to poor encryption. Data security is highly compromised. Authentication vulnerabilities are the gateway to unauthorized access and thus advance in identity theft. DoS attacks may overload resources, causing disruption of services. Some hardware vulnerabilities, like physical tampering, come with some of the highest risks; they include management and secure updates.

*Vulnerabilities in PaaS (Platform as a Service)*

Wrong network configuration puts the infrastructure at the risk of unauthorized access. There is a need for secure protocols and vulnerability assessment. The vulnerabilities related to code execution can compromise the tools of the platform, as well as the applications that are deployed on the platform. Likewise, the vulnerabilities in the apps can soil the integrity of the application. Data vulnerabilities due to insecure practices and weak API security can result in a breach and fines. IAM vulnerabilities expose platform services and sensitive data, and misconfigurations dangerously imperil platform services and APIs. This, in turn, contributes to waste. Access control vulnerabilities could make sensitive data fall into the wrong hands; cryptographic vulnerabilities would make insecure data exposures. Authentication vulnerabilities allow unauthorized access, leading to identity spoofing and breaches.

*Vulnerabilities in SaaS (Software as a Service)*

Code execution vulnerabilities compromise data and functionality, and application vulnerabilities lead to service disruption and information leakage. Data-related vulnerabilities are incurred due to insecure configurations and by not using multi-factor authentication, which breaches and compromises privacy. In fact, IAM vulnerabilities expose data to unauthorized access, leading to overall breaches and functional disruptions. Misconfigurations expose data and affect execution, leading to leaks. Access control vulnerabilities, such as unauthorized access to data and functionalities, would only lead to a breach. Similarly, cryptographic vulnerabilities would lead to access that is unauthorized and thereby a loss of privacy, while authentication vulnerabilities would lead to unauthorized access to applications and data. DoS vulnerabilities would render the application unavailable Hardware vulnerabilities would render the SaaS accessible to tampering and failures: and trusted vendors are essential to have strong security on devices.

*Vulnerabilities in Management and Governance*

The unscrupulous can exploit network management interface weaknesses to gain unauthorized access and assume control. The execution code vulnerabilities of such interfaces might easily lead to unauthorized changes in working configurations, which can really disrupt services. The real impact of application security issues is misconfigurations in the setup, which ultimately affect the entire security posture. This would mean that sensitive configuration data and other credentials are exposed thereby tempting unauthorized modifications. Inadequacy in IAM could lead unauthorized people to access settings and tools through undermining available security measures. Malformed configuration may result in undesired modification causing disruption of services. Inadequate control access mechanisms do provide loopholes to unauthorized users through exploiting such, causing risks directives. Some of the cryptographic protocols might highlight flaws and end up revealing certain sensitive information, thereby impacting governance. Vulnerabilities in authentication might grant access control to unauthorized individuals. Denial of Service (DoS) vulnerabilities may interfere with the operation of management services as well as monitoring. Hardware vulnerabilities are among the most pressing threats to cloud management services, indicating the significance of having proper security mechanisms for unauthorized access and Data Distributions.

# IV. NPM Audit

In the backdrop of vendor specific vulnerabilities and their effect on the cloud, the need to drive down to the specific project-based vulnerabilities are of high stake.

Specialized tools for software supply chain security are crucial in safeguarding the DevSecOps pipeline against attacks on the supply chain.Apart from risk detection, these tools enforce security protocols and ensure compliance across the software development lifecycle (SDLC) by scrutinizing each component, especially those obtained from external sources [27].

Software development teams benefit significantly from utilizing Software Composition Analysis (SCA) tools to meticulously monitor and evaluate any open-source code integrated into a project from the perspectives of licensing compliance and security vulnerabilities. These tools are adept at identifying open-source code, both superficially and in- depth, along with their direct and indirect dependencies, applicable licenses, and the presence of any documented security vulnerabilities or potential threats[28]. In general, these tools are deployed during development and build phases to detect vulnerabilities coming from third-party sources.

Recently, NPM security has garnered considerable attention in the media, with a notable focus on NPM packages within the ecosystem rather than the NPM registry itself.

Academic research published on June 7, 2019, shared interesting insights about this topic. Upon installing an average NPM package, users inadvertently place trust in 79 third-party packages and 39 maintainers, significantly enlarging the potential attack surface [29]. Researchers at ReversingLabs discovered that more than two dozen NPM packages, some of which dated back to at least December 2021, included code meant to collect form data from end users of the applications or websites that were hosting the malicious packages. Researchers do not know the full scope of the assault, but the malicious packages are presumably used by hundreds of downstream applications and websites[30].

Upon closer inspection, evidences have been discovered of a coordinated supply chain attack, with a large number of NPM packages containing jQuery scripts designed to steal form data from deployed applications that include them [30].

Attackers can choose from several methods to infect open source packages and distribute malicious code[31]

A security audit involves evaluating package dependencies to identify security vulnerabilities, thereby safeguarding users from risks such as data loss, service disruptions, unauthorized access to sensitive data, and other related issues. The NPM- audit feature is a crucial component of the NPM (Node Package Manager) toolkit, specifically designed for Node.js to conduct security audits on package dependencies.

## A. NPM Audit: Empowering Developer Awareness and Security

The NPM audit is one of the major activities that help developers become aware of any known security vulnerabilities in their applications. It provides actionable guidance to eliminate the identified risks, thus helping developers take immediate actions to ward off any potential threats.

By reading information in the project's `package.json` and `package-lock.json` files, NPM-audit compares each declared dependency against databases of known vulnerabilities, such as the national vulnerability database maintained by the npm registry. This approach provides a means for developers to have their application dependencies free from known vulnerabilities [32].

Using NPM audit, it provides detailed reports to the developer about vulnerabilities that exist within dependencies. It rates these vulnerabilities further by severity, stating clearly remediation steps, like upgrading to a safe version, patching, or removing entirely vulnerable packages. This feedback enables the developers to address security issues immediately, hence reducing the chances of exploitation [33]. In case of no vulnerabilities, NPM returns exit code 0.

One of the major advantages of NPM audit is the network effect. The more developers run the tool and fix their packages, the more secure the overall situation of the NPM ecosystem will become.

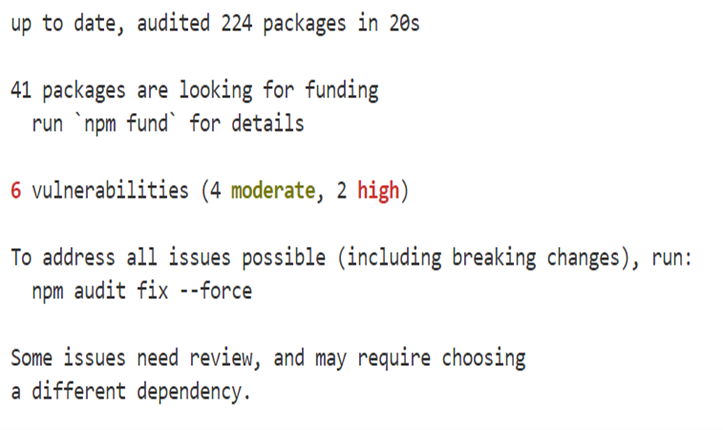
However, the visual analysis of the NPM audit data in detecting underlying patterns, enhancing understandability among a larger client base, etc., are some features that this tool has not yet implemented under its umbrella of functionalities.

A npm audit visualization dashboard requires the proper detection of the presence of packages with the highest count of vulnerabilities, their classification in accordance with common weakness enumeration, and the severity distribution assessment, among others. Provision of a collocated NPM Audit Visualization Dashboard that will dynamically render such details based on the vulnerabilities checked in the uploaded package.json file is the aim.

## B. Overview of Functionality

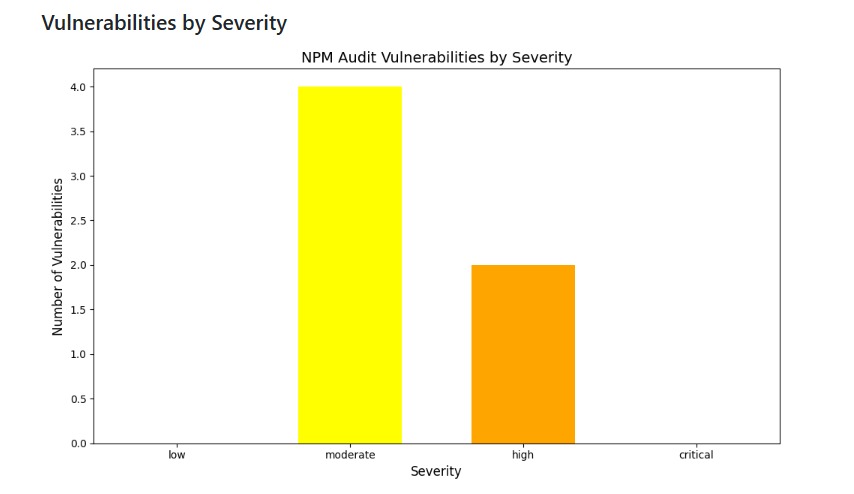
The base of this visualization is NPM, a package manager that scans the package.json file in real-time. The uploading of the package.json file connected with the project will trigger this NPM audit feature, and such a process will generate an audit dataset. By further parsing this dataset, key-value pairs corresponding to several parameters can be obtained, which will then be used as the base for visualizations, thus giving full insight into the state of security related to dependencies to developers.

It will hence provide a better, more informed, and quantified decision and a comparison before adding any new package to the development pipelines.



**Fig. 7.** Command for Prompt

### 1) Vulnerability Severity Analysis

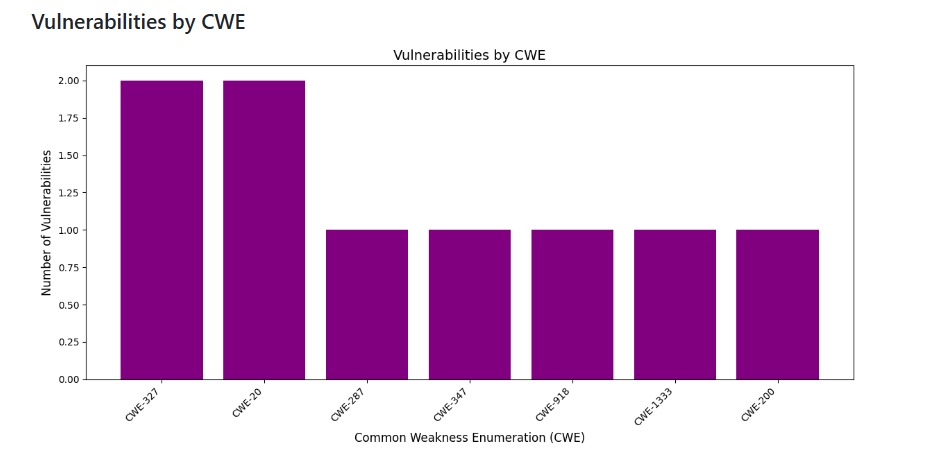


**Fig. 8.** Graph for Classification of Vulnerabilities by Severity

*Graph Explanation:* The chart classifies the vulnerabilities into four classes of severity: low, moderate, high, and critical.

*Technical Specifications:* NPM audit describes vulnerabilities with a severity score. The graph consolidates the scoring to help in prioritizing remedial action. The company can start by addressing critical and high-severity vulnerabilities to minimize major risks so that this will keep the development environment of the project safe from any threats.

### 3) Classification According to CWE

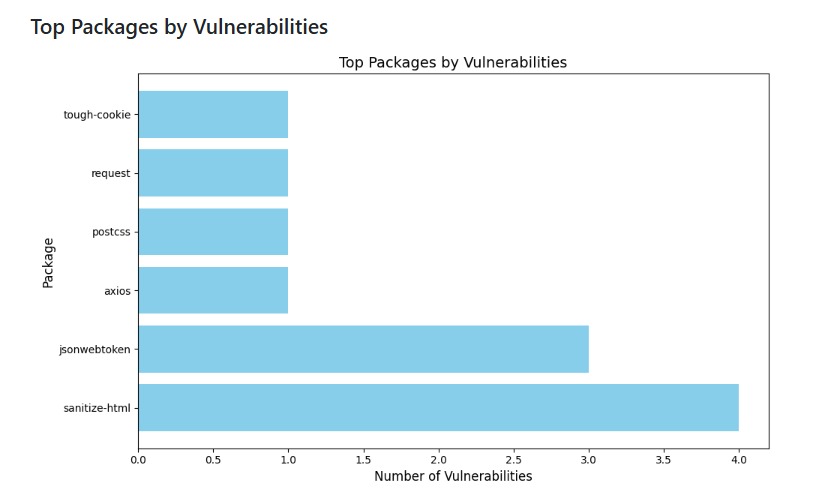


**Fig. 10.** Graph for classification according to CWE

*Graph Explanation:* This classifier then classifies these detected vulnerabilities based on the CWE framework.

*Technical Details:* This facilitates the categorization of vulnerabilities by CWE categories, letting developers understand the risk exposure of their dependencies. A developer might, in such a case, make use of already developed detection and mitigation strategies against an issue if a vulnerability either falls into one of the top-10 CWEs or any other high-risk category. This gives a clear understanding not only of the nature of the vulnerabilities but also eases remediation with very specific and more relevant solutions.

### 2) Package Vulnerability Analysis

**Fig. 9.** Graph for Top Packages by Vulnerabilities

*Graph Explanation:* This visualization displays which packages have been subject to the largest number of vulnerabilities.

*Technical Specifications:* Developers can use this to identify the weakest links amongst third-party/open-source packages by analysing the frequency of vulnerabilities in different packages. Understanding this forms a very critical basis while dealing with supply chain attacks and associated software vulnerabilities that can have high impacts on project confidentiality and financial stability.

### 4) Implementation Details

*Real-Time Data:* The visualizations are driven by dynamic information extracted from real-time results of npm audit, thus providing the most current security data to developers.

*Parsing of Audit Dataset:* Parsing of the audit dataset made available by npm audit is done for extracting details such as package names, severity levels, and CWE identifiers. This parsed information is further used in generating visualizations.

*Visual Representations:* There are already a number of important graphs on the dashboard, but many more can be implemented. All such visualizations are built to be user-friendly, insightful, and developer-friendly.

### 5) Advantages

*Efficient Use of Resources:* Knowing how the severity is distributed would help companies use their resources efficiently by first tackling the most critical vulnerabilities.

*Identification of Vulnerable Components:* The package analysis graph assists developers in recognizing and addressing the most vulnerable dependencies, enhancing the overall security of the project.

*Risk Mitigation:* The CWE-based classifier provides a structured approach to vulnerability management, enabling developers to implement effective detection and mitigation strategies based on the type of vulnerabilities present.

# V. PREDICTIVE MODEL BASED ON MACHINE LEARNING

Machine learning models are algorithms that can detect patterns or make predictions from previously unknown datasets. Scikit-learn is a machine learning toolkit that includes tools for many elements of machine learning such as , Regression, Classification, , Dimensionality reduction, Clustering, Model selection, Preprocessing.

## A. Data Preprocessing

Data preprocessing is a very necessary step in the preparation of raw data for use within the machine learning model [34]. Its objective is to simplify the data complexities used in machine learning algorithms. Real-world-based data normally is missing, and inconsistent. These are due to human mistakes, unexpected events, faults, and many other factors [35].

### 1) Data Cleaning

Data cleaning is an intrinsic part of data science that ensures the information one is dealing with is correct and reliable. It involves handling missing values and checking data for any inconsistencies [36].

### 2) Data Integration

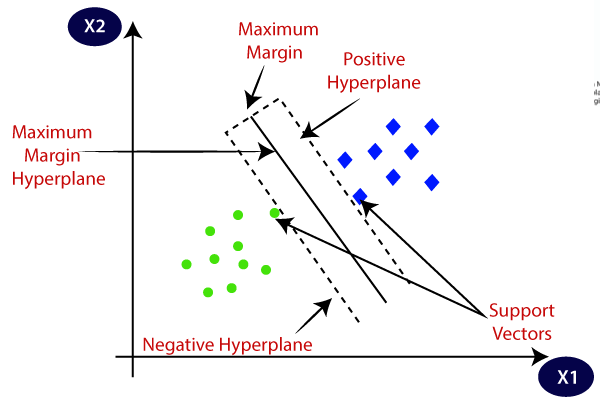
Data Integration is one of the data preprocessing steps, which involves integration of data and its subsequent combination into some larger database, for example, a Data Warehouse. This step becomes very important, especially when dealing with real-world problem-solving scenarios [37]. This becomes difficult to manage because of the variation in data format, structure, and meaning from one data source to another. Another data source may utilize different data categories, naming conventions, and data models, all of which may cause complications in an attempt to combine the data into a single view [38].

### 3) Data Transformation

Data transformation is the process of transforming unprocessed data into a format or structure that is better suited for analysis, to enhance its quality, and to align it with the needs of a specific task or systems[39].

### B. Mathematical analysis of the SVM algorithm

Support Vector Machine, also known as SVM, is one of the most widely used supervised learning algorithms. It can be used to solve regression and classification problems. However, in Machine Learning, it is primarily utilized for Classification issues. Support Vector Classifiers, or SVCs, are a kind of supervised machine learning algorithm that can be used for both regression and classification tasks. Due to their ability to achieve high accuracy while remaining computationally efficient, SVCs are a popular choice for many machine learning applications. SVMs function by first mapping data points to a high-dimensional space and then locating the most effective boundary between the various classes. The data points that are situated on either side of this boundary are referred to as support vectors. This boundary is referred to as a hyperplane. The support vectors determine the hyperplane's location. Data points that are closer to the hyperplane and have an effect on the hyperplane's position and orientation are called support vectors [40].The below image shows a hyperplane segregating two classes under the SVM model:



**Fig. 11.** SVM Graph

For the dataset used in this context, the kernel being set to linear gives highest precision and accuracy as the classification of the severity among the various classes does not involve very large levels of complexity and dispersion amongst themselves. Thus multiple linear hyperplanes can be used to classify the severity data effectively.

The severity output can fall into any one of the three classes- medium, high or critical. Thus, two linear hyperplanes are required for this classification.

Classes in the model: [‘CRITICAL’, ‘HIGH’, ‘MEDIUM’]

The hypothesis equation is defined as:

(1)

where:

* is the weight vector normal to the hyperplane, which decodes the orientation
* is the feature vector of the point
* is the bias term of the hyperplane

Any data point will either lie on the positive or the negative side of the hyperplane that separates the two classes. This point above or on the hyperplane will be classified as class +1, and the point below the hyperplane will be classified as class - 1. The hypothesis function helps to determine the same [41].

**Soft-margins:**

In a real time scenario, it might be possible that some data- points belonging to one hyperplane is placed on the incorrect side of the hyperplane or in other words is misclassified

The soft-margin SVM allows some points to be on the wrong side of the hyperplane but tries to keep the number and degree of such errors small. It does this by introducing a concept called slack variables ξ, which measure how much a point is on the wrong side. The goal of the soft-margin SVM is to find the best hyperplane by minimizing an objective function that balances two things: Maximizing the Margin and Minimizing Misclassification.

The mathematical expression to minimize is:

(2)

* is the term that encourages a larger margin (a simpler and more generalizable model).
* is the penalty term for misclassified points, where C is a parameter that controls the trade-off between margin size and misclassification penalty.

## D. Data Transformation of the vulnerability dataset

### 1) Encoding Categorical Values

Transforming categorical data into a numerical format is called encoding. It allows algorithms to straightforwardly identify and handle the data [42]. Encoding for CWE and Attack Vector which are taken as input parameters for the Machine Learning Model is done . Since both of these parameters have varied categories, Label Encoding is used for data transformation. Label Encoding is a technique that is used to convert categorical/descriptive columns into numerical ones[43].

### 2) Vectorization

Vectorization merely refers to the process of textual information being represented as numerical vectors. This is normally conducted after text cleaning. It might increase the speed of code execution and considerably reduce time used for training [44]. We used TF-IDF vectorization for the input parameter short description of the Machine Learning Model. This parameter gives a brief overview of the nature, type, cause and impact of the vulnerability.

TF-IDF is the shortened form of Term Frequency–Inverse Document Frequency, a statistical measure that expresses the importance of a word in a document. The calculation involves the number of times the word occurs divided by the total amount of words within that same document. Inverse Document Frequency, on the other hand, is a measure of the amount of information the word conveys [45].

The mathematical formula for Term Frequency (TF):

(3)

where *t* is a term (word) and *d* is a document

The mathematical formula for Inverse Document Frequency:

(4)

where:

* is the total number of documents in the corpus *D*.
* is the number of documents where the term *t* appears.
* The in the denominator is added to prevent division by zero if the term is not present in any document.

TF-IDF for a term in a document is computed as:

(5)

## D. Implementation of the SVC algorithm in context to supply chain vulnerability prediction and mitigation

**Algorithm: SVC Model for Vulnerability Classification**

**Input:** Dataset containing 'CWE category', 'Attack vector', and 'Short Description'

**Output:** Trained SVC model ,Mitigation strategies for each CWE category

*Start Algorithm*

Load and Clean Dataset removing noise.

Encode Features: Encode 'CWE category' and 'Attack vector' using Label Encoder. Encode 'Short Description' using TF-IDF Vectorizer

Split Dataset : Split dataset into training set (80%) and testing set (20%).Define features X: 'CWE encoded', 'Vector encoded', 'Description tfidf' and Define target y: 'Severity'

Train Model: Initialize SVC model with linear kernel and OvR strategy. Train model on training set X\_train , y\_train

Evaluate Model: Calculate accuracy on testing set X\_test, y\_test

Provide Mitigation Strategies: For each input CWE category: Retrieve CWE link and display mitigation strategies from CWE website

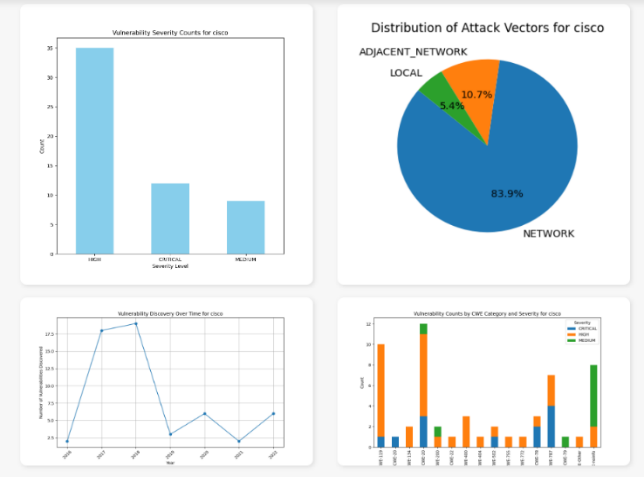
*End Algorithm*

# VI. Results and discussion

The npm audit visualization tool is a key tool in the analysis and representation of a project's vaguely structured and extended data on vulnerabilities. It improves the efficiency of the mitigation process by providing a developer with a clear and concise depiction of vulnerabilities.

The Severity Predictor is one of the better tools for analyzing vulnerability data in order to predict the potential severity of an attack. Such a targeted approach would ensure the implementation of the most effective security counters in line with the severity level anticipated, hence reducing the likelihood of successful attacks and reducing potential harm to system.

The Vulnerability Dashboard provides insight into deep current vulnerabilities faced by a specific third-party vendor. That is very useful and powerful information in many ways for the given vendor, the client of that vendor, and any Cloud Service Providers to be careful with forming appropriate security measures vis-à-vis the threatening factors related to vulnerability. Detailed dashboard reports enable proactive security defence strategies to ensure the management of any impending security threats before malicious entities have a chance to exploit them. Consider the vulnerability analysis of cisco-



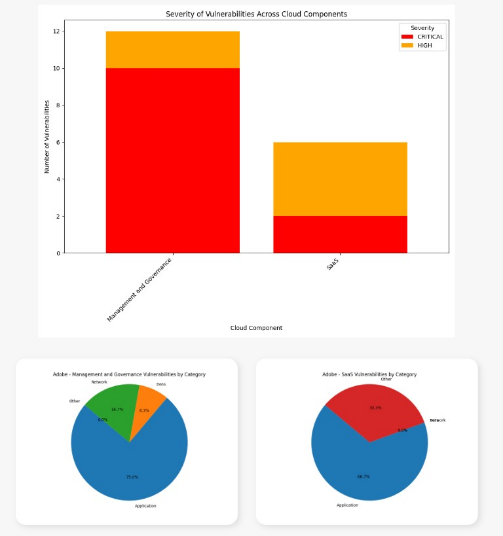
**Fig. 11.** Dashboard for Vulnerabilities

The vulnerability-based graphs provide us with a lot of vital information.

* The severity and attack vector graphs indicate that the majority of attacks propagate through networks rather than being localized.
* Most vulnerabilities are associated with either high or critical severity.
* The time-based graphs clearly depict a peak in the number of attacks in 2019, followed by a declining trend in subsequent years.
* The CWE classifier reveals that the majority of attacks fall under the CWE-20 category.

The cloud-based graphs classify which cloud component the attack is likely to effects based on the description and attack vector. For adobe most of the critical attacks affect the management and governance sector of the cloud infrastructure and within it specifically the application security domain.

The graphs for adobe are below



**Fig. 12.** Dashboard for Cloud Components

Cloud application security is the process of securing cloud-based software applications throughout the development lifecycle. It includes application-level policies, tools, technologies and rules to maintain visibility into all cloud-based assets, protect cloud-based applications from cyberattacks and limit access only to authorized users.

# VII. Future Works

The visualizations make planning, detection, and further analysis much easier to carry out and build upon. This cohesive implementation is a proper warning system for CSPs and their third-party vendors within their domains in advance, so that they can perform mitigation procedures to alleviate the effects and protect the computing systems from possible threats.

It would associate more effective strategies if tied to a **real- time detector of such weak links** by understanding the trends of vulnerabilities through our research in various packages and third-party vendors. This detector would focus on common attack entry points and more vulnerable dependencies, providing a more effective mitigation approach

# References

[1] Md J. H. Faruk, M. Tasnim, H. Shahriar, M. Valero, A. Rahman, and F. Wu, "Investigating Novel Approaches to Defend Software Supply Chain Attacks," in *2022 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW)*, 2022, pp. 283-283.

[2] "What is Software Vulnerability?", JFrog. [Online]. Available:<https://jfrog.com/devops-tools/article/software-vulnerability/>. [Accessed: Jul. 10, 2024].

[3] "PMC7338168," NCBI. [Online]. Available:<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7338168/>. [Accessed: Jul. 10, 2024].

[4]"Cloud Computing Supply Chain Attacks and the Open Source Effect," Voice&Data. [Online]. Available:<https://www.voicendata.com/cloud-computing-supply-chain-attacks-the-open-source-effect/>. [Accessed: Jul. 10, 2024].

[5] "Supply Chain Breaches Increased from 2022 to 2023," Supply Chain Brain. [Online].  Available:<https://www.supplychainbrain.com/articles/38672-supply-chain-breaches-increased-from-2022-to-2023>. [Accessed: Jul. 10, 2024].

[6] Snyk, "Software Supply Chain Report 2023," Snyk, 2023. [Online]. Available:  <https://go.snyk.io/2023-supply-chain-attacks-report.html>

[7] BlueVoyant, “The State of Supply Chain Defense: Annual Global Insights Report 2023”, 2023. [Online]. Available: <https://www2.bluevoyant.com/TheStateofSupplyChainDefense2023Report>

[8] "Most Supply Chain Attacks Target Suppliers’ Code," Trend Micro. [Online]. Available:<https://www.trendmicro.com/en_in/research/21/h/most-supply-chain-attacks-target-suppliers-code-enisa.html>. [Accessed: Jul. 10, 2024].

[9] "A Kill Chain Analysis of the 2013 Target Data Breach," Covert.io. [Online]. . [Accessed: Jul. 10, 2024].

Available:<https://www.covert.io/research-papers/security/A%20Kill%20Chain%20Analysis%20of%20the%202013%20Target%20Data%20Breach.pdf>

[10] "Software Supply Chain Attacks: Everything You Need to Know," PortSwigger. [Online]. Available:<https://portswigger.net/daily-swig/software-supply-chain-attacks-everything-you-need-to-know>. [Accessed: Jul. 10, 2024].

[11] "The Impact of Supply Chain Attacks on Cloud Services," Cloud Tech Insight. [Online]. Available:<https://cloudtechinsight.com/the-impact-of-supply-chain-attacks-on-cloud-services/#:~:text=In%20conclusion%2C%20the%20potential%20impact%20of%20a%20supply,reputation%2C%20regulatory%20compliance%20issues%2C%20and%20increased%20cybersecurity%20risks>. [Accessed: Jul. 10, 2024].

[12] "Assessing Cloud Risk: The Supply Chain Perspective," BCS. [Online]. Available:<https://www.bcs.org/articles-opinion-and-research/assessing-cloud-risk-the-supply-chain-perspective/>. [Accessed: Jul. 10, 2024].

[13] "Security Magazine - Companies Give Third Parties Access to All Cloud Data," Security Magazine. [Online]. Available:<https://www.securitymagazine.com/articles/94435-of-companies-give-third-parties-access-to-all-cloud-data#:~:text=The%20Wiz%20Research%20team%20conducted%20extensive%20research%20of,security%20risk%2C%20as%20well%20as%20serious%20privacy%20risk>.

[14] "6 Tips for Understanding 3rd Party Risk in the Cloud," Cloud Security Alliance. [Online]. Available:<https://cloudsecurityalliance.org/blog/2023/01/11/6-tips-for-understanding-3rd-party-risk-in-the-cloud>. [Accessed: Jul. 10, 2024].

[15] B. Gokkaya, L. Aniello, and B. Halak, \*Software Supply Chain: Review of Attacks, Risk Assessment Strategies, and Security Controls\*. 2023, pp. 1-27.

[16] A. Kruti, U. Butt, and R. Bin Sulaiman, \*A Review of SolarWinds Attack on Orion Platform Using Persistent Threat Agents and Techniques for Gaining Unauthorized Access\*. 2023, pp. 1-6.

[17] "Known Exploited Vulnerabilities," Panorays. [Online]. Available:<https://panorays.com/blog/known-exploited-vulnerabilities/>. [Accessed: Jul. 10, 2024].

[18] "Known Exploited Vulnerabilities Catalog," CISA. [Online]. Available:<https://www.cisa.gov/known-exploited-vulnerabilities-catalog>. [Accessed: Jul. 10, 2024].

[19] "Common Vulnerabilities and Exposures," CVE. [Online]. Available:<https://www.cve.org/>. [Accessed: Jul. 10, 2024].

[20] "Link Springer Article," Springer. [Online]. Available:<https://link.springer.com/article/10.1007/s10207-023-00751-6>. [Accessed: Jul. 10, 2024].

[21] "How to Prioritize Vulnerabilities," Paladin Cloud. [Online]. Available:<https://paladincloud.io/unified-vulnerability-management/how-to-prioritize-vulnerabilities/>. [Accessed: Jul. 10, 2024].

[22] "Infrastructure as a Service," Dot Net Tutorials. [Online]. Available:<https://dotnettutorials.net/lesson/infrastructure-as-a-service/>. [Accessed: Jul. 10, 2024].

[23] "Design Strategy for Securing IaaS, PaaS, and SaaS," Whizlabs. [Online]. Available:<https://www.whizlabs.com/blog/design-strategy-securing-iaas-paas-and-saas/>. [Accessed: Jul. 10, 2024].

[24] "SaaS, PaaS, and IaaS: What are All the Risks?," Arrow. [Online]. [Accessed: Jul. 10, 2024]. Available:<https://www.arrow.com/globalecs/na/arrow-channel-advisor/saas-paas-and-iaas-what-are-all-the-risks/>.

[25] "PaaS Security Threats, Solutions, and Best Practices," Spot.io. [Online]. Available:<https://spot.io/resources/cloud-security/paas-security-threats-solutions-and-best-practices/>. [Accessed: Jul. 10, 2024].

[26] "Cloud Security Architecture for IaaS, PaaS, and SaaS," BlueXP. [Online]. Available:<https://bluexp.netapp.com/blog/blg-cloud-security-architecture-for-iaas-paas-and-saas>. [Accessed: Jul. 10, 2024].

[27] "Software Supply Chain Security Tools," Snyk. [Online]. Available:<https://snyk.io/series/software-supply-chain-security/supply-chain-security-tools/>. [Accessed: Jul. 10, 2024].

[28] "An Open Guide to Evaluating Software Composition Analysis Tools," Linux Foundation Japan. [Online]. Available:<https://www.linuxfoundation.jp/wp-content/uploads/2020/11/An-Open-Guide-To-Evaluating-Software-Composition-Analysis-Tools_V2.pdf>. [Accessed: Jul. 10, 2024].

[29] NPM Security: Preventing Supply Chain Attacks," Snyk. [Online]. Available:<https://snyk.io/blog/npm-security-preventing-supply-chain-attacks/>. [Accessed: Jul. 10, 2024].

[30] "Dozens of Malicious Data Harvesting NPM Packages Found," Duo. [Online]. [Accessed: Jul. 10, 2024].Available:  [https://duo.com/decipher/dozens-of-malicious-data-harvesting-npm-packages-found#:~:text=Researchers%20have%20uncovered%20a%20software%20supply-chain%20attack%20involving,were%20embedded%20in%20various%20mobile%20applications%20and%20websites](%20https://duo.com/decipher/dozens-of-malicious-data-harvesting-npm-packages-found#:~:text=Researchers%20have%20uncovered%20a%20software%20supply-chain%20attack%20involving,were%20embedded%20in%20various%20mobile%20applications%20and%20websites).

[31] "Data Study Reveals Predictors of Supply Chain Attacks in NPM Repositories," PortSwigger. [Online]. Available:<https://portswigger.net/daily-swig/data-study-reveals-predictors-of-supply-chain-attacks-in-npm-repositories>. [Accessed: Jul. 10, 2024].

[32] "OSV Scanner vs. npm Audit: A Detailed Comparison of SCA Tools," Jit.io. [Online]. Available: <https://www.jit.io/resources/appsec-tools/osv-scanner-vs-npm-audit-a-detailed-comparison-of-sca-tools>. [Accessed: Jul. 10, 2024].

[33] "The Basics of Package.json," NodeSource. [Online]. Available: <https://nodesource.com/blog/the-basics-of-package-json>. [Accessed: Jul. 10, 2024].

[34] "Data Preprocessing in Machine Learning," Simplilearn. [Online]. Available: <https://www.simplilearn.com/data-preprocessing-in-machine-learning-article>.[Accessed: Jul. 10, 2024].

[35] "Data Preprocessing Guide," Neptune.ai. [Online]. Available: <https://neptune.ai/blog/data-preprocessing-guide> [Accessed: Jul. 10, 2024].

[36] "Data Preprocessing in Machine Learning using Python," GeeksforGeeks. [Online]. [Accessed: Jul. 10, 2024].Available:<https://www.geeksforgeeks.org/data-preprocessing-machine-learning-python/>

[37] "Data Preprocessing Guide," V7 Labs. [Online]. Available: <https://www.v7labs.com/blog/data-preprocessing-guide> [Accessed: Jul. 10, 2024].

[38] "Data Integration in Data Mining," GeeksforGeeks. [Online]. Available: <https://www.geeksforgeeks.org/data-integration-in-data-mining/> [Accessed: Jul. 10, 2024].

[39] "Data Transformation in Machine Learning," GeeksforGeeks. [Online]. Available: <https://www.geeksforgeeks.org/data-transformation-in-machine-learning/> [Accessed: Jul. 10, 2024].

[40] "Support Vector Machine Algorithm," JavaTpoint. [Online]. Available: <https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm>.[Accessed: Jul. 10, 2024].

[41] "Understanding the Mathematics Behind Support Vector Machines," Fritz.ai. [Online]. Available: <https://fritz.ai/understanding-the-mathematics-behind-support-vector-machines/> [Accessed: Jul. 10, 2024].

[42] "Encoding Techniques for Machine Learning in Alteryx: A Concise Guide," Alteryx Community. [Online]. Available: <https://community.alteryx.com/t5/Data-Science/Encoding-Techniques-for-Machine-Learning-in-Alteryx-A-Concise/ba-p/1203894> [Accessed: Jul. 10, 2024].

[43] "Categorical Data," DataCamp. [Online]. Available: <https://www.datacamp.com/tutorial/categorical-data>

 [Accessed: Jul. 10, 2024].

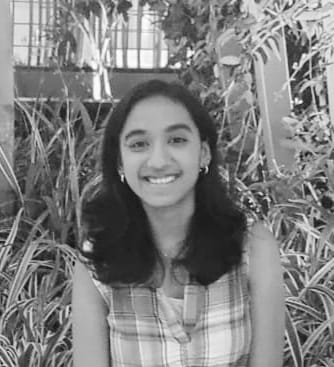
[44] "Vectorization in Machine Learning," Heartbeat by Comet. [Online]. Available: <https://heartbeat.comet.ml/vectorization-in-machine-learning-2e3bdce7dbe> [Accessed: Jul. 10, 2024].

[45] "Vectorization Techniques in NLP: A Guide," Neptune.ai. [Online]. Available: <https://neptune.ai/blog/vectorization-techniques-in-nlp-guide>

[Accessed: Jul. 10, 2024].

[46] "TF-IDF Vectorizer Scikit-Learn," Medium. [Online]. Available: <https://medium.com/@cmukesh8688/tf-idf-vectorizer-scikit-learn-dbc0244a911a>[Accessed: Jul. 10, 2024].

**Varshini M** is a Student in PES University currently pursuing Bachelors in Computer Science and Engineering. She is interested in fields comprising Cloud Computing and Data Analytics. She is currently associated with the Research Center for Information Security, Digital Forensics and Cyber Resilience, PES University.

**Srishti Mathur** is a Student in PES University currently pursuing Bachelors in Computer Science and Engineering. She is interested in fields comprising Cloud Computing and Networks, and is currently associated with the Research Center for Information Security, Digital Forensics and Cyber Resilience, PES University.

 **Nagasundari S** received Ph.D degree in the faculty of Information and Communication Engineering from Anna University, India. She is currently working as an Associate Professor in the Department of Computer Science and associated with the Research Center for Information Security, Digital Forensics and Cyber Resilience and Research Center for Internet of Things, PES University. She is a member of ACM and IEEE. She is also a member in Automotive Security Research Group (ASRG), Institution of Engineers India (IEI) and Cloud Security alliance (CSA), Bangalore Chapter. Her research interests include Cloud Security and IoT Security.

1. [↑](#footnote-ref-1)