Article

Generative Pre-Trained Transformers, Natural Language Processing and Artificial Intelligence and Machine Learning (AI/ML) in Software Vulnerability Management: automations in the Software Bill of Materials (SBOM) and the Vulnerability-Exploitability eXchange (VEX)

Petar Radanliev * , David De Roure , Omar Santos

Posted Date: 19 July 2023

doi: 10.20944/preprints202307.1303.v1

Keywords: Artificial Intelligence and Machine Learning (AI/ML); Cyber vulnerability management; Software Bill of Materials (SBOM); Vulnerability-Exploitability eXchange (VEX); Common Security Advisory Framework (CSAF); Software Supply Chain Cyber Risk

Article

Generative Pre-Trained Transformers, Natural Language Processing and Artificial Intelligence and Machine Learning (AI/ML) in Software Vulnerability Management: Automations in the Software Bill of Materials (SBOM) and the Vulnerability-Exploitability eXchange (VEX)

Petar Radanliev [1,*], David De Roure [1] and Omar Santos [2]

1 Department of Engineering Science, University of Oxford, david.deroure@oerc.ox.ac.uk
2 Cisco Systems, RTP, North Carolina, United States, email; osantos@cisco.com
* Correspondence: petar.radanliev@eng.ox.ac.uk

Abstract: One of the most burning topics in cybersecurity in 2023 will undoubtedly be the compliance with the **Software Bill of Materials.** Since the US president issued the Executive Order 14028 on Improving the Nation's Cybersecurity, software developers have prepared and bills are transmitted to vendors, customers, and users, but they don't know what to do with the reports they are getting. In addition, since software developers have identified the values of the **Software Bill of Materials**, they have been using the reports extensively. This article presents an estimate of 270 million requests per month, just from form one popular tool to one vulnerability index. This number is expected to double every year and a half. This simple estimate explains the urgency for automating the process. We propose solutions based on artificial intelligence and machine learning, and we base our tools on the existing FAIR principles (Findable, Accessible, Interoperable, and Reusable). This methodology is supported with a case study research and Grounded theory, for categorising data into axis, and for verifying the values of the tools with experts in the field. We showcase how to create, and share **Vulnerability Exploitability eXchange** data, and automate the **Software Bill of Materials** compliance process with AI models and a unified computational framework combining solutions for the following problems: (1) the data utilisation problem, (2) the automation and scaling problem, (3) the naming problem, (4) the alignment problem, (5) the pedigree, and provenance problem, and many other problems that are on the top of mind for many security engineers at present. The uptake of these findings will depend on collaborations with government and industry, and on the availability and the ease of use of automated tools.

Keywords: artificial intelligence and machine learning (AI/ML); cyber vulnerability management; software bill of materials (SBOM); vulnerability-exploitability eXchange (VEX); common security advisory framework (CSAF); software supply chain cyber risk

1. Introduction

This article addresses problems associated with cyber vulnerability management, going into details about **Common Security Advisory Framework** (CSAF), and the relationship between the **Software Bill of Materials** (SBOM) and the **Vulnerability Exploitability eXchange** (VEX) and the implementation. The main aim of the article is to present solution for automation in the SBOM/VEX implementation process, because asset owners operate with priorities related to revenue, uptime, maintenance, but cybersecurity is not one of their key priorities. Regulatory requirements can be priority, especially if they are related to safety, asset owners will have to comply with the regulation, but it's not their main priority for operations. Since cybersecurity is not a priority, it's also quite down on the list of operations. This article is covering only a small part of the cybersecurity operations, which is the updates and patch management. Narrowing the focus even more, from the three different types of patches (1) feature update, (2) bug fixes, and (3) security fixes, the topic of interest in this article is most closely related to security fixes. The reason is that for the first two patches, there is a routine maintenance schedule, but for the security fixes, it is difficult to allocate any other time than the time of resolution.

1.1. Data Sources

Although the number of records was very limited, the results were extremely interesting. We discovered that VEX has been considered for use in the US Nuclear Industry [6], for resolving software supply chain insecurities in vehicles [7], and the most interesting document was a Bachelors thesis on the 'Development of an API to request security advisories for CSAF 2.0' [8]. The document is surprisingly detailed, and given more time, this effort could have solved the CBOM/VEX integration and automation problem. Bachelors' thesis is however a very short project, and as a rule, 8-12 weeks are allowed maximum. Nevertheless, the thesis presents a solid grounding on building the API proposed in this paper. The Google search on "software bill of materials" resulted with 367 records, and the most relevant records are included in the case study research as secondary data sources on specific issues, but there is no solution to the SBOM/VEX automation problem in current literature.

1.2. Methods for Analysis

To analyse the data records, the case study research method is applied, in combination with grounded theory to categorise data records into axes. The emerging solutions are derived from multiple sources and verified through workshops with the study participants. Experimental testing of the solutions was conducted in the Oxford e-Research Centre.

2. Brief Literature Rreview of Cybersecurity Reports on VEX and SBOM

2.1. What is Software Supply Chain Cyber Risk

Software Supply Chain can be defined as the collection of components, libraries, tools, and processes used to develop, build, and publish the software. In supply chain manufacturing, there is a well-established concept of a 'bill of materials' (BOM) [9]. In more recent developments, the same principles are applied to software supply chains. The 'Cyber Supply Chain Management and Transparency Act of 2014' [10] proposed that US government agencies obtain SBOMs for all new software. This led to the 'Internet of Things Cybersecurity Improvement Act of 2017' [11], and more recently, 'The US Executive Order on Improving the Nation's Cybersecurity of May 12, 2021 [12] ordered The National Institute of Standards and Technology (NIST) to issue guidance on '*providing a purchaser a Software Bill of Materials (SBOM) for each product.'*

2.2. What Is the Software Bill of Materials (SBOM) and How Does It Help with Cyber Risk Assessment

The **Software Bill of Materials** (SBOM) can be defined as a nested (machine-readable) inventory for software, a list of ingredients that make up software components and dependencies, and their hierarchical relationships [13,14]. The main use cases include supply chain assets and vulnerabilities management via sharing and exchanging SBOMs, but because of '*the diverse needs of the software ecosystem, there is no one-size-fits-all solution'* [15]. The problem with sharing and exchanging is that *'To fully realize the benefits of SBOMs and software component transparency, machine processing and automation are necessary'* [15]. In **Error! Reference source not found.**, we can see the hierarchy of automation in different formats (SPDX, CycloneDX, and SWID), specifications, and tools that are still under development.

2.3. Background to the Vulnerability Management Problems

SBOM enables us to identify potentially vulnerable components, but a vulnerability associated with a software component is not necessarily exploitable. The current CVE index has over 191633 vulnerabilities [16], with 32760 new vulnerabilities published in 2022, and around 22,000 published in 2021. Organisations cannot perform cybersecurity risk management for all known vulnerabilities, and risk management is based on their cyber risk tolerance, which is based on the likelihood and severity (frequency and magnitude) of the risk materialising. If a vulnerability is not assessed for exploitability, it is impossible to predict likelihood. Just one Windows vulnerability opened the door for WannaCry ransomware attacks via the EternalBlue exploit, while the Mirai botnet that spread through the exploitation of multiple vulnerabilities. From this, we can conclude that all vulnerabilities need to be risk assessed, and according to CVE, almost 11% of all vulnerabilities can be categorised as 'critical' [17] because they enable hackers to compromise apps and data of users of the same hardware. In early 2020, the first critical vulnerabilities in a major cloud infrastructure were discovered - disproving '*the assumption that cloud infrastructures are secure*' [18]. Vulnerabilities also need to be risk assessed because around 75% of attacks in 2020 used vulnerabilities were at least two years old [19] and high-risk vulnerabilities are present on the network perimeters of 84% of companies [20].

2.4. What Is the Vulnerability Exploitability eXchange (VEX) and How Is It Different from SBOM and CVE

The examples above explain why each vulnerability needs to be assessed for exploits. From the volume of vulnerabilities in the CVE index, we can easily understand why the risk assessment process needs to be automated. From the most severe vulnerability of 2021, the Log4j, we can understand why vulnerabilities need to be assessed for exploits – while most cybersecurity professionals wasted vast amount of time risk assessing if Log4j was exploitable on their system, in many cases, it was not exploitable, and yet, it is extremely likely that a high number of Log4j vulnerable applications remain online [21]. To help prevent this in the future, the **Vulnerability Exploitability eXchange** (VEX)[22] was created in 2021. VEX's providess the SBOM's with transparency and an up-to-date view of the status of vulnerabilities. Software suppliers can issue a VEX to prevent non-exploitable vulnerabilities being investigated. VEX has been implemented as a profile in the Common Security Advisory Framework (CSAF) [23].

3. Survey of Secondary Data on Rethinking VEX and SBOM

Utilisation Problem

At present, SBOMs are used extensively by software developers, and given how useful SBOMs have been for them, developers are very keen on distributing the reports to end users (i.e., organisations where the primary business is not software development). The problem is that end users are not using SBOMs in any meaningful way, and when they get the reports, they don't know what to do with them. Furthermore, without transparency, and without the end users requesting for SBOMs, it is quite possible that most of the software developers are not even using SBOMs to secure their software products. Without a reporting mechanism for the utilisation of SBOMs, we simply cannot know. However, from discussions with software developers on various meetings and conferences (e.g., with CSAF, NTIA, CISA), we know that developers have been very successful in using SBOMs, but that is not commonly known by end users. The motivation for this article, is not to determine the value of SBOMs in securing end users' networks, that topic has been covered extensively in year 2021/22. Determining the value of SBOMs is not a primary or major obstacle at present. The motivation for this article is to determine what will persuade end users to start using SBOMs, and the key to this is to determine **what is the real value** in SBOMs for end users. One specific use case is the automated use of SBOMs for CVE (vulnerability) management. Although there are few tools designed to manage the software development process of vulnerability management, what is missing is '*easy-to-use and low cost tools and third party services… passing vulnerability data to the vulnerability and configuration management tools that are now deployed by end users – is currently being addressed by nobody*' [26]. The most useful tool that could be found in open source at present (that ingests, analyses, monitors, and produces real-time intelligence reports from SBOMs and VEX documents), is the Dependency-Track [27].

4. Conclusions

Although CSAF has made some significant contributions towards the improvement of machine readable SBOM and VEX records, which is a crucial step in the automation process, NIST already has all 200k CVE descriptions available for download in JSON format (nvd.nist.gov). Without designing the second phase of the automation process, we struggle to see the rationale for CSAF. The CSAF Security Advisories (profile 4) are essential to helping consumers manage software risk using automation. The main difference between the CSAF and current mitigation strategies, is that the CSAF is machine-readable and allows for rapid and automatic mitigation.

The new version of CSAF released in December 2022 is more than just a JSON file readable in a machine format. The standard specifies how the new CSAF documents, and their distribution system can be discovered. The VEX profile in CSAF addresses the main problem that emerged with the introduction of SBOMs, it provides negative advisories that can be inform the customer that a specific product version they use and is listed in a SBOM, might not be affected at all by a certain vulnerability, because it might have been patched, resolved or simply the system is not affected.

The main benefit of CSAF is the ability to respond to false positives from security scanners. Although not well known in the European cybersecurity circles, the most recent CSAF standard version is stable and already in use by Cisco, Oracle, Arista, Siemens, Red Hat, BSI, and other organisations.

Areas for Further Research

The OASIS Standard is the language reference for the CSAF version 2.0 and is designed to automate the exchange of Security Advisories formulated in JSON. NFT is a JSON string. Future research should investigate if the security of product naming can be solved by naming software products as non-fungible tokens (NFTs). Generative AI will also provide many solutions to the product naming and different file formats problems. Generative Pre-Trained Transformers can be personalised and specialised to be specific for the VEX problems, and prevent the reports being disseminated for vulnerabilities that should have been patched yesterday, which seems to be the case with manual patching.

Declarations:

Availability of data and materials: all data and materials are included in the article

Funding: This work has been supported by the PETRAS National Centre of Excellence for IoT Systems Cybersecurity, which has been funded by the UK EPSRC [under grant number EP/S035362/1], the Software Sustainability Institute [Grant Number:], and by the Cisco Research Centre [grant number CG1525381].

Acknowledgements: Eternal gratitude to the Fulbright Visiting Scholar Project.

References

[1] Wiesner, Jens, "CSAF, Not SBOM, Is The Solution," *S4x22 - BSI*, 2022. [Online]. Available: https://www.youtube.com/watch?v=fKlW9vOs7X4&t=504s. [Accessed: 03-Jan-2023], URL: https://www.youtube.com/watch?v=fKlW9vOs7X4&t=504s.

[2] NIST, "NVD - CVSS v3 Calculator," *CVSS Version 3.1*, 2022. [Online]. Available: https://nvd.nist.gov/vuln-metrics/cvss/v3-calculator. [Accessed: 03-Jan-2023], URL: https://nvd.nist.gov/vuln-metrics/cvss/v3-calculator.

[3] Manion, Art, "SSVC: Stakeholder-Specific Vulnerability Categorization," *Carnegie Mellon University*, 2020. [Online]. Available: https://bit.ly/3ambIP4. [Accessed: 02-Jan-2023], URL: https://bit.ly/3ambIP4.

[4] Carmody, Seth., Coravos, Andrea., Fahs, Ginny., Hatch, Audra., Medina, Janine., Woods, Beau., and Corman, Joshua, "Building resilient medical technology supply chains with a software bill of materials," *npj Digital Medicine 2021 4:1*, vol. 4, no. 1, pp. 1–6, Feb. 2021, doi: 10.1038/s41746-021-00403-w.

doi:10.20944/preprints202307.1303.v1

[5] Foster, Rita., Priest, Zach., and Cutshaw, Michael, "Infrastructure eXpression for Codified Cyber Attack Surfaces and Automated Applicability," *2021 Resilience Week, RWS 2021 - Proceedings*, 2021, doi: 10.1109/RWS52686.2021.9611807.

[6] Eggers, Shannon Leigh., Christensen, Drew., Simon, Tori Brooke., Morgan, Baleigh Rae., and Bauer, Ethan S, "Towards Software Bill of Materials in the Nuclear Industry," Sep. 2022, doi: 10.2172/1901825.

[7] Moore, Marina., Sirish, Aditya., Yelgundhalli, A., Kuppusamy, Trishank Karthik., Torres-Arias, Santiago., Delong, Lois Anne., and Cappos, Justin, "Scudo: A Proposal for Resolving Software Supply Chain Insecurities in Vehicles The second in a series of whitepapers on emerging and critical issues in automotive cybersecurity."

[8] Schmidt, Leon., Hammer, Daniel., Biß, Klaus., and Schmidt, Thomas, "Media Faculty in cooperation with the Development of an API to request security advisories for CSAF 2.0 by Enterprise and IT Security Supervision."

[9] Jackson, Wayne, "Code, Cars, and Congress: A Time for Cyber Supply Chain Management (1 of 3)," *Sonatype*, 2014. [Online]. Available: https://blog.sonatype.com/2014/12/cyber-supply-chain-management-part1/. [Accessed: 03-Jan-2023], URL: https://blog.sonatype.com/2014/12/cyber-supply-chain-management-part1/.

[10] Royce, Edward R., "H.R.5793 - 113th Congress (2013-2014): Cyber Supply Chain Management and Transparency Act of 2014," *Congress.Gov*, 2014. [Online]. Available: http://www.congress.gov/. [Accessed: 03-Jan-2023], URL: http://www.congress.gov/.

[11] Howard, Matt, "Cybersecurity Improvement Act of 2017: The Ghost of Congress Past - DevOps.com," *devops.com*, 2017. [Online]. Available: https://devops.com/cybersecurity-improvement-act-2017-ghost-congress-past/. [Accessed: 03-Jan-2023], URL: https://devops.com/cybersecurity-improvement-act-2017-ghost-congress-past/.

[12] Biden, Joseph, "Executive Order on Improving the Nation's Cybersecurity | The White House," *The White House*, 12-May-2021. [Online]. Available: https://www.whitehouse.gov/briefing-room/presidential-actions/2021/05/12/executive-order-on-improving-the-nations-cybersecurity/. [Accessed: 03-Jan-2023], URL: https://www.whitehouse.gov/briefing-room/presidential-actions/2021/05/12/executive-order-on-improving-the-nations-cybersecurity/.

[13] CISA, "Software Bill of Materials," *Cybersecurity & Infrastructure Security Agency*, 2018. [Online]. Available: https://www.cisa.gov/sbom. [Accessed: 24-Dec-2022], URL: https://www.cisa.gov/sbom.

[14] NTIA, National Telecommunications and Information Administration, *Software Bill of Materials (SBOM) | National Telecommunications and Information Administration*. 2021.

[15] NTIA, "SBOM at a Glance," *NTIA Multistakeholder Process on Software Component Transparency | ntia.gov/sbom*, 2021. [Online]. Available: https://tiny.cc/SPDX. [Accessed: 03-Jan-2023], URL: https://tiny.cc/SPDX.

[16] MITRE, "CVE - Common Vulnerabilities and Exposures," *The MITRE Corporation*, 2022. [Online]. Available: https://cve.mitre.org/. [Accessed: 03-Jan-2023], URL: https://cve.mitre.org/.

[17] CVE, "CVE security vulnerability database. Security vulnerabilities, exploits, references and more," 2022. [Online]. Available: https://www.cvedetails.com/. [Accessed: 03-Jan-2023], URL: https://www.cvedetails.com/.

[18] Shustin, Ronen, "Remote Cloud Execution – Critical Vulnerabilities in Azure Cloud Infrastructure (Part II) – Check Point Research," *Check Point Research*, 2020. [Online]. Available: https://research.checkpoint.com/2020/remote-cloud-execution-critical-vulnerabilities-in-azure-cloud-infrastructure-part-ii/. [Accessed: 03-Jan-2023], URL: https://research.checkpoint.com/2020/remote-cloud-execution-critical-vulnerabilities-in-azure-cloud-infrastructure-part-ii/.

[19] CheckPoint, "Cyber Security Report," *Check Point*, 2021. .

[20] Postitive Techologies, "Vulnerabilities on the corporate network perimeter," 2020. [Online]. Available: https://www.ptsecurity.com/ww-en/analytics/vulnerabilities-corporate-networks-2020/. [Accessed: 03-Jan-2023], URL: https://www.ptsecurity.com/ww-en/analytics/vulnerabilities-corporate-networks-2020/.

[21] O'Driscoll, Aimee, "25+ Cyber Security Vulnerability Statistics and Facts of 2023," *Comparitech*, 2022. [Online]. Available: https://www.comparitech.com/blog/information-security/cybersecurity-vulnerability-statistics/. [Accessed: 03-Jan-2023], URL: https://www.comparitech.com/blog/information-security/cybersecurity-vulnerability-statistics/.

[22] NTIA, the U.S. National Telecommunications and Information Administration, "Vulnerability-Exploitability eXchange (VEX)," 2021, URL: https://ntia.gov/files/ntia/publications/vex_one-page_summary.pdf.

[23] OASIS, "OASIS Common Security Advisory Framework (CSAF) TC | OASIS," *OASIS OPEN*, 2022. [Online]. Available: https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=csaf. [Accessed: 03-Jan-2023], URL: https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=csaf.

[24] Springett, Steve, "Vulnerability and Exploitability Transparency - VDR & VEX | OWASP Foundation," *OWASP*, Feb-2023. [Online]. Available: https://owasp.org/blog/2023/02/07/vdr-vex-comparison. [Accessed: 24-Apr-2023], URL: https://owasp.org/blog/2023/02/07/vdr-vex-comparison.
[25] CISA, "Minimum Requirements for Vulnerability Exploitability eXchange (VEX)," *U.S. Department of Homeland Security, Cybersecurity and Infrastructure Security Agency*, 2023, URL: http://www.cisa.gov/tlp/.
[26] Alrich, Tom, "'Minimum elements', Bigfoot, and other myths," *Blog*, 2022. [Online]. Available: https://tomalrichblog.blogspot.com/2022/. [Accessed: 03-Jan-2023], URL: https://tomalrichblog.blogspot.com/2022/.
[27] Dependency-Track, "Software Bill of Materials (SBOM) Analysis | OWASP," *Dependency-Track*, 2022. [Online]. Available: https://dependencytrack.org/. [Accessed: 03-Jan-2023], URL: https://dependencytrack.org/.
[28] OSS, "Sonatype OSS Index," 2022. [Online]. Available: https://ossindex.sonatype.org/. [Accessed: 03-Jan-2023], URL: https://ossindex.sonatype.org/.
[29] Alrich, Tom, "Hmm…It seems SBOMs might become big one of these days…," *Blog*, 2022. [Online]. Available: https://tomalrichblog.blogspot.com/2022/04/hmmit-seems-sboms-might-become-big-one.html. [Accessed: 03-Jan-2023], URL: https://tomalrichblog.blogspot.com/2022/04/hmmit-seems-sboms-might-become-big-one.html.
[30] NTIA, Multistakeholder Process on Software Component Transparency - Standards and Formats Working Group, "Survey of Existing SBOM Formats and Standards-Version 2021 Survey of Existing SBOM Formats and Standards Credit: Photo by Patrick Tomasso on Unsplash NTIA Multistakeholder Process on Software Component Transparency Standards and Formats Working Group," 2021, URL: https://www.ntia.gov/files/ntia/publications/sbom_formats_survey-version-2021.pdf.
[31] VEX, "Vulnerability Exploitability eXchange (VEX) – Use Cases," 2022. [Online]. Available: https://www.cisa.gov/sbom. [Accessed: 03-Jan-2023], URL: https://www.cisa.gov/sbom.
[32] Alrich, Tom., and Brudo, Barak, "Secrets Of The Vulnerability Exploitability eXchange (VEX) Webinar," *Scribe Securely*, 2022. [Online]. Available: https://www.youtube.com/watch?v=dyvuNqX_wJ8. [Accessed: 03-Jan-2023], URL: https://www.youtube.com/watch?v=dyvuNqX_wJ8.
[33] Wilkinson, Mark D., Dumontier, Michel., Sansone, Susanna Assunta., Bonino da Silva Santos, Luiz Olavo., Prieto, Mario., Batista, Dominique., McQuilton, Peter., … Schultes, Erik, "Evaluating FAIR maturity through a scalable, automated, community-governed framework," *Scientific Data 2019 6:1*, vol. 6, no. 1, pp. 1–12, Sep. 2019, doi: 10.1038/s41597-019-0184-5.
[34] Groll, Elias.,., and Hewitt-Jones, John, "Software bills of materials face long road to adoption," *CyberScoop*, 2022. [Online]. Available: https://www.cyberscoop.com/dhs-sbom-adoption/. [Accessed: 03-Jan-2023], URL: https://www.cyberscoop.com/dhs-sbom-adoption/.
[35] Meyers, John Speed, "Are SBOMs Any Good? Preliminary Measurement of the Quality of Open Source Project SBOMs," *Chainguard*, 2022. [Online]. Available: https://www.chainguard.dev/unchained/are-sboms-any-good-preliminary-measurement-of-the-quality-of-open-source-project-sboms. [Accessed: 03-Jan-2023], URL: https://www.chainguard.dev/unchained/are-sboms-any-good-preliminary-measurement-of-the-quality-of-open-source-project-sboms.
[36] Alrich, Tom, "Real-time VEX," *Blog*, 07-Sep-2022. [Online]. Available: https://tomalrichblog.blogspot.com/2022/09/real-time-vex.html. [Accessed: 03-Jan-2023], URL: https://tomalrichblog.blogspot.com/2022/09/real-time-vex.html.
[37] CISA, "CISA Stakeholder-Specific Vulnerability Categorization Guide," *Cybersecurity and Infrastructure Security Agency*, Nov-2022. .
[38] OASIS, "Using CSAF to Respond to Supply Chain Vulnerabilities at Large Scale," *OASIS Open*, 2022. [Online]. Available: https://us06web.zoom.us/webinar/register/WN_KqD-a1t5SpuMI7w9cI7ZDg. [Accessed: 03-Jan-2023], URL: https://us06web.zoom.us/webinar/register/WN_KqD-a1t5SpuMI7w9cI7ZDg.
[39] Wilkinson, Mark D., Dumontier, Michel., Aalbersberg, IJsbrand Jan., Appleton, Gabrielle., Axton, Myles., Baak, Arie., Blomberg, Niklas., … Mons, Barend, "The FAIR Guiding Principles for scientific data management and stewardship," *Scientific Data 2016 3:1*, vol. 3, no. 1, pp. 1–9, Mar. 2016, doi: 10.1038/sdata.2016.18.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.