CIFRE Inria - Teclib : CVE-GLPI

“Automated detection and classification of machines containing vulnerable softwares in an Information System”

Keywords

Vulnerabilities, CVE, CPE, inventory of IT assets.

Context

A vulnerability is a defect of a software that offers to a potential attacker possibilities to exploit and compromise an IT system, targeting spying, destruction or ransom extortion. Considering only ransomware, 90% of financial institutions [1,2] have been targeted by ransomware attacks, and more than 68000 new trojan ransomware for mobiles have been discovered in 2019 [1,3]. Consequences of these attacks are very important: their cost has reached more than 7.5 billions of dollars in 2019 [1,4]; average downtime of a company after a ransomware attack was 21 days [1,5] in 2020 fourth quarter. And 80% of victim companies that have paid a ransom have suffered another attack shortly after. The problem of attacks and vulnerabilities is therefore a major issue.

A vulnerability has a life cycle, starting with “zero-day” and ending in a corrected vulnerability. Vulnerabilities are disclosed via official channels such as National Vulnerability Database or github Security Advisories which publish about one hundred vulnerabilities every day.

An IT system management solution such as GLPI [5,6], developed by Teclib, allows to gather all the details on installed software and their versions for all machines constituting the IT system of a company. A tool, using these inventory data to evaluate which versions of installed software are vulnerable and to which vulnerability, would be a major help for the defense against cyber-attacks.

Subject

The work of this PhD thesis in computer science will consist in studying fundamental and practical issues in designing a system that will match machines inside an IT system with vulnerabilities affecting software installed on these machines, in order to signal to system administrators the vulnerability level and the need to protect or update each machine.

To achieve this goal, the work will include the following steps:

1) Study how to build the inventory of software, of their dependencies and of the versions installed in an IT system, whilst taking into account the diversity of installations: operating systems (Windows, Linux, Mac OS), open-source or proprietary software, source code available in public repositories or not available, software installed using package managers or using build systems [10-14]. Several agents may be used or developed in order to perform these tasks.
2) Study catalogs of disclosed vulnerabilities (CVE [8] or others) and the existing means to match them with software and versions: explicit and formalized links (such a CPE [9]) or informal links (such as natural language text describing the vulnerability).
3) Extract appropriate informations from system inventory and vulnerabilities catalogs in order to match installed software with known vulnerabilities affecting them

4) Provide to system administrators the list of machines affected by vulnerabilities, with explicit confidence/certitude levels, and signal the level of severity/urgency/importance of the corrections to be applied.

In all these steps, the modularity of the approach will be emphasized, in order to be able to address the various technical problems in a progressive, replaceable and extensible way.

Some issues to be solved: system reliability with a very low false positive/negative rate, high scalability allowing to analyze quickly (about every hour) 10000 machines with 1000 software installed on each machine.

Eventually, the system will be integrated into the open source solution GLPI developed by Teclib and will therefore be distributed under an open source license. The integration of this system will have a high impact on the marketing of GLPI by adding a major advantage against competitive solutions.

References

[1] 81 Ransomware Statistics, Data, Trends and Facts for 2021.
https://www.varonis.com/blog/ransomware-statistics-2021
[2] Financial institutions; 90% of them have been targeted by ransomware.
https://www.prdistribution.com/news/financial-institutions-90-of-them-have-been-targeted-by-ransomware/329096
[3] 20 Ransomware Statistics You're Powerless to Resist Reading.
https://www.thesslstore.com/blog/ransomware-statistics/
[4] The State of Ransomware in the US: Report and Statistics 2019.
https://blog.emsisoft.com/en/34822/the-state-of-ransomware-in-the-us-report-and-statistics-2019/
[5] Coveware Quarterly Ransomware Report Q4 2020.
https://www.coveware.com/blog/ransomware-marketplace-report-q4-2020
[6] Gestionnaire Libre de Parc Informatique.
https://fr.wikipedia.org/wiki/Gestionnaire_Libre_de_Parc_Informatique
[7] Gestion des services informatiques.
https://fr.wikipedia.org/wiki/Gestion_des_services_informatiques
[8] Common Vulnerabilities and Exposures
https://www.cve.org/
[9] Common Platform Enumeration
https://nvd.nist.gov/products/cpe
[10] S. Peisert, B. Schneier, H. Okhravi, F. Massacci, T. Benzel, C. Landwehr, M. Mannan, J. Mirkovic, A. Prakash, and J. B. Michael, "Perspectives on the solarwinds incident," IEEE Security Privacy, vol. 19, no. 2, pp. 7–13, 2021.
[11] T. Herr, "Breaking trust–shades of crisis across an insecure software supply chain," 2021.
[12] H. Assal and S. Chiasson, "Security in the software development lifecycle," in Fourteenth symposium on usable privacy and security (SOUPS 2018), pp. 281–296, 2018.
[13] B. A. Sabbagh and S. Kowalski, "A socio-technical framework for threat modeling a software supply chain," IEEE Security Privacy, vol. 13, no. 4, pp. 30–39, 2015.
[14] D.-L. Vu, "Typosquatting and combosquatting attacks on the python ecosystem," 07 2020.

Environment / Supervisory

CIFRE: thesis will be financed by a CIFRE grant (Cifre | Association Nationale Recherche Technologie)

Teclib: Teclib is a small company developing open source software for enterprises. Teclib's offer is built around GLPI (Gestion Libre de Parc Informatique), an open source ITSM solution. GLPI provides standard functionalities such as helpdesk, knowledge base, administrative and financial reporting. GLPI provides an inventory tool that allows to build a detailed inventory of the machines in an IT system; this inventory contains for each machine data such as components, disk space, installed software and versions of these software… GLPI is marketed by Teclib and an international partners network, as the GLPI-Network professional distribution that offers support and assistance. A SAAS offer of GLPI is also marketed.

DiverSE (Inria): DiverSE is an Inria research team in software engineering, with applications in cybersecurity. Our observation is that the required flexibility and openness raise challenges for the software layer of these systems that must deal with four dimension of diversity: the diversity of languages used by the stakeholders involved in the construction of these systems, the diversity of features (aka variability) required by the different customers, the diversity of runtime environments in which software has to run and adapt, and the diversity of implementations that are necessary for resilience through redundancy. We study the production and delivery of modern software systems that involve the integration of diverse dependencies and continuous deployment on diverse execution platforms in the form of large distributed socio-technical systems. This leads to new software architectures and programming models, as well as com-plex supply chains for final delivery to system users. In order to boost cybersecurity, we want to provide strong support to software engineers and IT teams in the development and deployment of secure and resilient software systems, ie. systems able to resist or recover from cyberattacks. DiverSE has expertise in software build (e.g., large-scale build, incremental build), coevolution, and cybersecurity.

Duration: 3 years

Place: Thesis will take place mainly at INRIA-Rennes site (Beaulieu scientific campus), inside the DiverSE team, and also inside Teclib (Caen,...).

Profile

Master in computer science or equivalent.

Good knowledge in development and software engineering.

Good level in english.

Curiosity, motivation, autonomy, ability to work inside a team, abstraction capabilities, programming ability, interest for open source.

Knowledge in automatic classification and machine learning is appreciated.

Contacts

Olivier Barais (Olivier.Barais@inria.fr), Olivier Zendra (Olivier.Zendra@inria.fr), François Déchelle (fdechelle@teclib.com).