CIFRE Inria - Teclib : CVE-GLPI

“Détection et classification automatisée et fiable de machines comportant des logiciels vulnérables dans un parc informatique”

Mots-clés

Vulnérabilités, CVE, CPE, inventaire d’actifs informatiques,

Contexte

Une vulnérabilité est une malfaçon d’un logiciel, offrant à un attaquant potentiel des possibilités d’exploitation et de compromission de systèmes informatiques, à des fins d'espionnage, de destruction, ou d’extorsion de rançons. Si on regarde les rançongiciels (*ransomware*) seuls, 90 % des institutions financières [1,2] ont été ciblées par des attaques de rançongiciels, et plus de 68 000 nouveaux chevaux de Troie rançongiciels pour mobile ont été découverts en 2019 [1,3]. Les conséquences de ces attaques sont majeures : leur coût a dépassé les 7,5 milliards de dollars en 2019 [1,4]; le temps d'arrêt moyen que subit une entreprise après une attaque par rançongiciel était de 21 jours [1,5] au quatrième trimestre 2020. Et 80 % des entreprises victimes qui ont payé une rançon ont subi une autre attaque peu de temps après [1,2]. Le problème des attaques et des vulnérabilités qui les facilitent est donc majeur.

Une vulnérabilité a un cycle de vie, de “zero-day” à vulnérabilité corrigée. Les vulnérabilités sont rendues publiques via des canaux officiels (National Vulnerability Database, github Security Advisories…) qui publient chaque jour de l’ordre de la centaine de vulnérabilités.

Un logiciel de gestion de parc informatique tel que GLPI [5,6], édité par Teclib, permet de connaître de façon détaillée l’ensemble des logiciels et de leurs versions installés sur l’ensemble de machines constituant le parc informatique d’une entreprise. Un outil évaluant, à partir de ces informations, quelles versions des logiciels installés sont vulnérables et à quelles vulnérabilités serait une aide majeure dans la lutte contre les attaques.

Sujet

Le travail de cette thèse de doctorat en informatique consistera à étudier les problèmes fondamentaux et pratiques posés par la conception d’un système permettant de mettre en correspondance les machines d’un parc informatique et les vulnérabilités qui sont présentes dans les logiciels installés sur ces machines, afin de pouvoir signaler aux administrateurs du parc le niveau de vulnérabilité et la nécessité de protéger (mettre à jour) chaque machine.

Pour ce faire, le travail consistera en les étapes suivantes.

1) Étudier les moyens permettant d’obtenir l'inventaire des logiciels, leurs dépendances et leurs versions présents sur un parc, c'est-à-dire sur chaque machine, en tenant compte de la variabilité des installations : système d'exploitation (windows, Linux, Mac OS), logiciels open-source ou propriétaires, code source accessible dans des dépôts publics ou code inaccessible, logiciels installés via des gestionnaires de packages ou via des systèmes de build [10-14]. Différents agents peuvent avoir à être utilisés ou conçus ex-nihilo pour réaliser ces tâches.
2) Étudier les catalogues de vulnérabilités existants (CVE [8] ou autres) et les moyens existants qui les lient aux logiciels et leurs versions : liens assez explicites et formels (par ex. via CPE [9]), ou liens informels (par ex. via le texte en langage naturel qui décrit la vulnérabilité).

3) Extraire les informations appropriées de l'inventaire du parc et des catalogues de vulnérabilités pour pouvoir mettre en correspondance les logiciels d'une machine et les vulnérabilités connues qu'ils contiennent.
4) Fournir au gestionnaire de parc informatique, avec des niveaux de confiance/certitudes explicités, la liste des machines affectées par les vulnérabilités, et indiquer le degré de gravité/urgence/importance des traitements à effectuer sur chaque machine.

Pour ces différentes étapes, la modularité de l'approche sera importante, afin de pouvoir attaquer les différents problèmes techniques de façon progressive, remplaçable, et extensible dans le futur.

Quelques verrous à lever : fiabilité du système produisant un très faible taux de faux positifs/négatifs, capacité de montée en charge permettant l'analyse rapide (périodicité d'environ 1 heure) de l'ordre de 10000 machines comportant chacune de l'ordre de 1000 logiciels/paquets installés.

À terme, ce système sera intégré dans la solution open source GLPI développée par Teclib et sera donc distribué sous une licence open source. L'intégration de ce système aura un impact fort pour la valorisation de GLPI en lui offrant un avantage concurrentiel majeur.

Références

[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.

Environnement / Encadrement

CIFRE: la thèse sera financée par le dispositif CIFRE (Cifre | Association Nationale Recherche Technologie)

Teclib: Teclib est une PME éditeur de logiciels open source pour les entreprises. L’offre de Teclib s’articule autour de GLPI (Gestion Libre de Parc Informatique), une solution ITSM open source. GLPI fournit des fonctionnalités standards telles que gestion d’assistance (“helpdesk”) , base de connaissances, suivi administratif et financier. GLPI intègre un outil d’inventaire permettant de construire un inventaire détaillé des machines composant le parc informatique géré par GLPI; cet inventaire contient pour chaque machine des informations telles que composants matériels, espace disque, logiciels installés et versions de ces logiciels… GLPI est commercialisé par Teclib et un réseau international de partenaire via la distribution professionnelle GLPI-Network qui comprend une offre de support et d’assistance. Une offre SAAS de GLPi est également commercialisée.

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 complex 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.

Durée: 3 ans, début prévu en septembre 2022.

Lieu: La thèse aura lieu en majorité sur le site de l’INRIA-Rennes équipe DiverSE (Campus scientifique Beaulieu à Rennes) et également au sein de la société Teclib (Caen,...)

Profil recherché

Master en informatique ou équivalent.

Connaissances en développement et en génie logiciel.

Bon niveau en anglais.

Curiosité, motivation, dynamisme, autonomie, capacité à travailler en équipe, capacité d’abstraction, capacité à programmer, intérêt pour l’open-source.

Connaissances en classification et apprentissage machine appréciées.

Contacts

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