Identification automatique des vulnérabilités de sécurité dans les systèmes logiciels

Raounak Benabidallah*[1]

[1]Institut de Recherche en Informatique et Systèmes Aléatoires (IRISA) – Universite de Rennes 1, Université de Rennes, Institut National des Sciences Appliquées - Rennes, Institut National des Sciences Appliquées, Université de Bretagne Sud, École normale supérieure - Rennes, Institut National de Recherche en Informatique et en Automatique, CentraleSupélec : UMR6074, Centre National de la Recherche Scientifique, IMT Atlantique Bretagne-Pays de la Loire, Institut Mines-Télécom [Paris] – Avenue du général LeclercCampus de Beaulieu 35042 RENNES CEDEX, France

Résumé

La menace posée par les vulnérabilités logicielles croît de manière exponentielle. Ce phénomène est dû, d'une part, à l'omniprésence des logiciels, et d'autre part, au nombre important de failles existantes. Pour faire face à ce problème, plusieurs stratégies ont été élaborées au fil du temps. Certaines visent à mettre en place de bonnes pratiques de développement et les intégrer dès la phase de conception tandis que d'autres consistent à effectuer des inspections de sécurité en indiquant les zones vulnérables. Les travaux présentés s'inscrivent dans la deuxième catégorie et portent essentiellement sur la construction de modèles de prédiction de vulnérabilités. La création de ces derniers soulève différents problèmes. Le plus important étant le manque de données sur les vulnérabilités logicielles. À cet effet, nous mettons en place une chaîne de traitement complète allant de la création et l'annotation automatique d'un corpus de sécurité jusqu'à la construction et l'évaluation des modèles de prédiction de vulnérabilités. La première contribution est plus axée sur l'approche de construction de corpus que sur le corpus lui-même. L'approche est basée sur la conception de méta-scanners de vulnérabilités permettant d'identifier des vulnérabilités de code efficacement. Cela consiste à combiner plusieurs outils d'analyse statique en se basant sur leurs performances individuelles pour chaque catégorie de vulnérabilités. Notre deuxième contribution correspond au corpus SecureQualitas qui consiste en un corpus d'applications Java annotées avec les vulnérabilités qu'elles contiennent. Nous construisons ce corpus en utilisant un méta-scanner construit à l'aide de trois outils d'analyse de vulnérabilités. Enfin, notre troisième contribution est de construire un modèle de prédiction du code vulnérable. Nous avons opté pour l'utilisation de métriques de qualité pour caractériser le code et nous avons étudié les performances des modèles à la fois sur des catégories de vulnérabilités apprises par les modèles et sur des catégories non encore connues Les résultats de nos expérimentations ont montré l'efficacité des modèles sur les deux populations de vulnérabilités : connues et non connues.

*Intervenant