Synthèse de Code pour la Déobfuscation de Binaires

Mots clés: Analyse de binaires, déobfuscation, intelligence artificielle, synthèse de code

Encadrant Principal
Grégoire MENGUY
CEA LIST
gregoire.menguy@cea.fr

Superviseur Sénior
Sébastien BARDIN
CEA LIST
sebastien.bardin@cea.fr

Un programme peu contenir de nombreux secrets comme des clés privées ou des algorithmes propriétaires. Un attaquant peu ainsi les voler en rétro-ingéniant le code. L'*obfuscation* propose de protéger les secrets contenus dans le programme en empêchant sa rétro-ingénierie. Elle transforme le programme original en un programme équivalent mais beaucoup plus difficile à comprendre. A l'inverse, la *déobfuscation* tente de simplifier un programme obfusqué pour retrouver une version proche de l'originale. L' évolution des méthodes de déobfuscation a ainsi de nombreuses implications en sécurité. D'un coté, cela permet de mieux estimer la robustesse des protections existantes, améliorant la sécurité des programmes. De l'autre, elle aide à analyser les logiciels malveillants, usuellement obfusqués.

Récemment, les méthodes en boîte noire [4], se sont révélées très efficaces, permettant de simplifier des programmes obfusqués avec des obfuscateurs complexes comme VMProtect ou Tigress. Ces approches boîtes noires se reposent sur de la synthèse de programme [1], et notamment sur des algorithmes de recherche locale. En parallèle, la communauté de synthèse a proposé différentes évolutions pour gérer des programmes plus complexes comme le *CEGIS* [3] ou l'utilisation de *term graph* [2].

Ce projet propose d'étudier ces différentes avancés et de les intégrer dans le déobfuscateur Xyntia développé au CEA. Il se déroulera comme suit:

- Faire une état de l'art des dernières avancées en synthèse de programme;
- Choisir une approche et l'implémenter dans le déobfuscateur open source Xyntia [4];
- Évaluer ces extensions sur du code obfusqué et sur des problèmes de synthèse de programme.

1 Livrables attendus

Les principaux livrables attendus sont:

- Un résumé des recherches bibliographiques menées;
- Une implémentation documentée de l'extension proposée;
- Le rapport final (rédigé ou sous forme de slides) incluant les deux premiers livrables et le récapitulatif des démarches suivis et des résultats obtenus.

2 Éléments logistiques

Des points réguliers seront organisés avec l'encadrement. Ils pourront être réalisés en visioconférence ou au CEA sur le site Nano-Innov à Saclay.

3 Postuler

Pour postuler, les étudiants doivent contacter l'encadrant par mail (avec le superviseur sénior en CC).

References

[1] Rajeev Alur et al. "Syntax-guided synthesis". In: *Formal Methods in Computer-Aided Design, FMCAD 2013, Portland, OR, USA, October 20-23, 2013*. IEEE, 2013.

[2] Yuantian Ding and Xiaokang Qiu. "Enhanced enumeration techniques for syntax-guided synthesis of bit-vector manipulations". In: *Proceedings of the ACM on Programming Languages* 8.POPL (2024), pp. 2129–2159.

[3] Priyanka Golia, Subhajit Roy, and Kuldeep S Meel. "Manthan: A data-driven approach for boolean function synthesis". In: *International Conference on Computer Aided Verification*. Springer. 2020, pp. 611–633.

[4] Grégoire Menguy et al. "Search-Based Local Black-Box Deobfuscation: Understand, Improve and Mitigate". In: *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*. CCS '21. Virtual Event, Republic of Korea: Association for Computing Machinery, 2021, pp. 2513–2525. ISBN: 9781450384544. DOI: 10.1145/3460120.3485250. URL: https://doi.org/10.1145/3460120.3485250.