

Proposta aiMalware: artificial intelligence and **Malware**

Introdução

Esta proposta consiste numa investigação, conceptualização e respetivo desenvolvimento experimental dum ferramenta de deteção de malware usando 3 tipos de técnicas de classificação: i) classificadores treinados a partir de conhecidas bases de dados de malware transformadas em datasets, ii) heurísticas exploratórias como por exemplo algoritmos genéticos para identificar situações de malware, e iii) heurísticas dedicadas à deteção de similaridades no código (por exemplo heurísticas baseadas em Locality Sensitive Hashing (LSH), Locality-preserving hashing (LPH)). Os 3 tipos de classificadores serão juntos usando uma técnica de Adaboost.

Tarefas a Realizar

1. Criar uma Taxonomia para o malware (setembro)
2. Estado da arte. Procura de bases de dados de malware e criação de um dataset de malware ou procurar em repositórios datasets de malware já existentes. Estudar algoritmos exploratórios de dados aplicados à detecção de malware (exemplo algoritmos genéticos, e ADA Análise Exploratória de Dados). Estudar técnicas de detecção de similaridade de conteúdos (do código) (exemplo LSH e LPH) (outubro-novembro)
3. Desenho de uma heurística exploratória para identificar situações de malware com implementação como prova de conceito (dezembro)
4. Desenho de uma heurística em busca de situações de similaridade no código com implementação como prova de conceito (janeiro)
5. Criação de um dataset de malware. Desenvolvimento de ferramentas de treino e construção de uma machine learning (fevereiro, março)
6. Combinação dos classificadores num classificador adaboost (abril)
7. Testes (maio)

Referências

- Sharma, S., Krishna, C.R. and Sahay, S.K., 2019. Detection of advanced malware by machine learning techniques. In Soft Computing: Theories and Applications (pp. 333-342). Springer, Singapore.
- Chumachenko, K., 2017. Machine Learning Methods for Malware Detection and Classification.
- Kim, C.W., 2018. NtMalDetect: a machine learning approach to malware detection using native API system calls. arXiv preprint arXiv:1802.05412.

Contactos

Orientador: Paul Crocker PhD, crocker@di.ubi.pt

Co-orientador: Paulo Vieira PhD, paulo.vieira@ubi.pt Cloud Computing Competence Centre, UBI