

VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
PROGRAMŲ SISTEMŲ STUDIJŲ PROGRAMA

GAN architektūrų, tinkamų kenkėjiško kodo obfuskacijai, analizė

Analysis of GAN architectures suitable for Ethical Malware Obfuscation

Kursinis darbas

Atliko: 4 kurso 3 grupės studentas

Liudas Kasperavičius

Darbo vadovas: prof. dr. Olga Kurasova

Vilnius – 2024

Turiny

NOTES	3
ĮVADAS	4
REZULTATAI IR IŠVADOS	5
ŠALTINIAI	6
SANTRUMPOS	8

Notes

- notes/1.md - Notes about [NDI⁺23]
- notes/2.md - Notes about [ZCY⁺24]
- notes/3.md - Notes about [ZHZ⁺22]
- notes/4.md - Notes about [KOD19]
- notes/5.md - Notes about [HT17]
- notes/6.md - Notes about [FWL⁺19]
- notes/7.md - Notes about [ZZY⁺22]
- notes/8.md - Notes about [CSD19]
- notes/9.md - Notes about [AKF⁺18]
- notes/10.md - Notes about [DCB⁺21]
- notes/11.md - Notes about [YPT22]
- notes/12.md - Notes about [CDH⁺16]
- notes/13.md - Notes about [DBL⁺21]
- notes/14.md - Notes about [RSR⁺18]

Įvadas

Pastaraisiais metais Dirbtinio intelekto (DI) naudojimas kenkėjiškų programų ar kenkėjiško kodo aptikimo srityje tapo itin populiarus **[CITATION NEEDED]**.

Tikslas – nustatyti labiausiai tinkantį modelį varžymosi principais pagrįstoms atakoms „juodos dėžės“ atvejais.

Uždaviniai:

1. Apžvelgti kenkėjiško kodo obfuskacijos metodus
2. Nustatyti kriterijus ir įvertinti modelius
3. Atlikti eksperimentinį tyrimą su nustatytu modeliu

Rezultatai ir išvados

Rezultatų ir išvadų dalyje turi būti aiškiai išdėstomi pagrindiniai darbo rezultatai (kažkas išanalizuota, kažkas sukurta, kažkas įdiegta) ir pateikiamos išvados (daromi nagrinėtų problemų sprendimo metodų palyginimai, teikiamos rekomendacijos, akcentuojamos naujovės).

Šaltiniai

- [AKF⁺18] H. S. Anderson, A. Kharkar, B. Filar, D. Evans, P. Roth. *Learning to Evade Static PE Machine Learning Malware Models via Reinforcement Learning*. 2018-01-30. [žiūrėta 2024-09-30]. Prieiga per internetą: <https://doi.org/10.48550/arXiv.1801.08917>.
- [CDH⁺16] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, P. Abbeel. *InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets*. 2016-06-11. [žiūrėta 2024-10-07]. Prieiga per internetą: <https://doi.org/10.48550/arXiv.1606.03657>.
- [CSD19] R. L. Castro, C. Schmitt, G. Dreo. AIMED: Evolving Malware with Genetic Programming to Evade Detection. Iš: *2019 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/13th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*. 2019, p. 240–247 [žiūrėta 2024-09-23]. ISSN 2324-9013. Prieiga per internetą: <https://doi.org/10.1109/TrustCom/BigDataSE.2019.00040>.
- [DBL⁺21] L. Demetrio, B. Biggio, G. Lagorio, F. Roli, A. Armando. Functionality-Preserving Black-Box Optimization of Adversarial Windows Malware. *IEEE Transactions on Information Forensics and Security*. 2021, tomas 16, p. 3469–3478 [žiūrėta 2024-10-14]. ISSN 1556-6021. Prieiga per internetą: <https://doi.org/10.1109/TIFS.2021.3082330>.
- [DCB⁺21] L. Demetrio, S. E. Coull, B. Biggio, G. Lagorio, A. Armando, F. Roli. Adversarial EXEmples: A Survey and Experimental Evaluation of Practical Attacks on Machine Learning for Windows Malware Detection. *ACM Trans. Priv. Secur.* 2021, tomas 24, numeris 4, 27:1–27:31 [žiūrėta 2024-09-30]. ISSN 2471-2566. Prieiga per internetą: <https://doi.org/10.1145/3473039>.
- [FWL⁺19] Z. Fang, J. Wang, B. Li, S. Wu, Y. Zhou, H. Huang. Evading Anti-Malware Engines With Deep Reinforcement Learning. *IEEE Access*. 2019, tomas 7, p. 48867–48879 [žiūrėta 2024-09-18]. ISSN 2169-3536. Prieiga per internetą: <https://doi.org/10.1109/ACCESS.2019.2908033>.
- [HT17] W. Hu, Y. Tan. *Generating Adversarial Malware Examples for Black-Box Attacks Based on GAN*. 2017-02-20. [žiūrėta 2024-09-18]. Prieiga per internetą: <http://arxiv.org/abs/1702.05983>.
- [YPT22] J. Yuste, E. G. Pardo, J. Tapiador. Optimization of Code Caves in Malware Binaries to Evade Machine Learning Detectors. *Computers & Security*. 2022, tomas 116, p. 102643 [žiūrėta 2024-10-07]. ISSN 0167-4048. Prieiga per internetą: <https://doi.org/10.1016/j.cose.2022.102643>.

- [KOD19] M. Kawai, K. Ota, M. Dong. Improved MalGAN: Avoiding Malware Detector by Learning Cleanware Features. Iš: *2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)*. 2019, p. 040–045 [žiūrėta 2024-09-14]. Prieiga per internetą: <https://doi.org/10.1109/ICAIIIC.2019.8669079>.
- [NDI⁺23] H. Nguyen, F. Di Troia, G. Ishigaki, M. Stamp. Generative Adversarial Networks and Image-Based Malware Classification. *Journal of Computer Virology and Hacking Techniques*. 2023, tomas 19, numeris 4, p. 579–595 [žiūrėta 2024-09-15]. ISSN 2263-8733. Prieiga per internetą: <https://doi.org/10.1007/s11416-023-00465-2>.
- [RSR⁺18] I. Rosenberg, A. Shabtai, L. Rokach, Y. Elovici. Generic Black-Box End-to-End Attack Against State of the Art API Call Based Malware Classifiers. Iš: M. Bailey, T. Holz, M. Stamatogiannakis, S. Ioannidis (sudarytojai). *Research in Attacks, Intrusions, and Defenses*. Cham: Springer International Publishing, 2018, p. 490–510. ISBN 978-3-030-00470-5. Prieiga per internetą: https://doi.org/10.1007/978-3-030-00470-5_23.
- [ZCY⁺24] F. Zhong, X. Cheng, D. Yu, B. Gong, S. Song, J. Yu. MalFox: Camouflaged Adversarial Malware Example Generation Based on Conv-GANs Against Black-Box Detectors. *IEEE Transactions on Computers*. 2024, tomas 73, numeris 4, p. 980–993 [žiūrėta 2024-09-15]. ISSN 1557-9956. Prieiga per internetą: <https://doi.org/10.1109/TC.2023.3236901>.
- [ZHZ⁺22] F. Zhong, P. Hu, G. Zhang, H. Li, X. Cheng. Reinforcement Learning Based Adversarial Malware Example Generation against Black-Box Detectors. *Computers & Security*. 2022, tomas 121, p. 102869 [žiūrėta 2024-09-14]. ISSN 0167-4048. Prieiga per internetą: <https://doi.org/10.1016/j.cose.2022.102869>.
- [ZZY⁺22] E. Zhu, J. Zhang, J. Yan, K. Chen, C. Gao. N-Gram MalGAN: Evading Machine Learning Detection via Feature n-Gram. *Digital Communications and Networks*. 2022, tomas 8, numeris 4, p. 485–491 [žiūrėta 2024-09-23]. ISSN 2352-8648. Prieiga per internetą: <https://doi.org/10.1016/j.dcan.2021.11.007>.

Santrumpos

DI. Dirbtinis Intelektas