Application of Reinforcement Learning to Creating Adversarial Malware Samples

Ing. Matouš Kozák, Supervisor: Mgr. Martin Jureček, Ph.D.

{matous.kozak,martin.jurecek}@fit.cvut.cz

Faculty of Information Technology, Czech Technical University in Prague

1. Abstract

- Adversarial machine learning in the malware detection domain is used to mislead antivirus programs.
- **We introduce a generator of adversarial malware samples based on reinforcement learning (RL) techniques.**
- The reinforcement learning agents are trained to modify binary malware executables.
- Our modifications are designed to better preserve the original functionality of the malware in comparison with other state-of-the-art RL-based generators.
- Using the proximal policy optimization (PPO) algorithm, we successfully avoided detection by gradient-boosted decision tree (GBDT) in 58.92% of cases.
- The same PPO agent previously trained against the GBDT classifier bypassed MalConv 28.91% of the time, a model based solely on machine learning (ML).
- **The adversarial samples generated by the PPO agent evaded detection by leading antiviruses in 10.24% to 25.7% of cases.**

2. Proposed Method

- We propose a generator called AMG (Adversarial Malware Generator), which works by iteratively perturbing input malware samples to avoid detection by the target classifier.
- AMG works in pure black-box settings, meaning no information apart from the labeled output of the target classifier is required.
- Our generator targets static malware analysis, i.e., malware classification based on features extracted from binaries without running them.
- Ten binary file modifications were implemented, such as appending benign content to overlay, adding new sections, or removing debug information.
- **Based on our functionality preservation testing, our modifications better preserve the original functionality than state-of-the-art generators such as gym-malware [1] or MAB-Malware [2].**

3. Experiments Description

- We tested three reinforcement learning algorithms: Deep Q-Network (DQN), Policy Gradients (PG), and PPO.
- We used a dataset consisting of 6,000 Windows malicious programs in the PE file format. The dataset was split into 4,000 training, 1,000 validation, and 1,000 testing samples.
- Firstly, we optimized the maximum number of allowed modifications and later tuned the learning rate (α) and discount rate (γ) hyperparameters.
- Lastly, the trained RL agents were tested against GBDT, MalConv, and commercially available antivirus (AV) products. The full overview of our experiments is depicted in Figure 1.

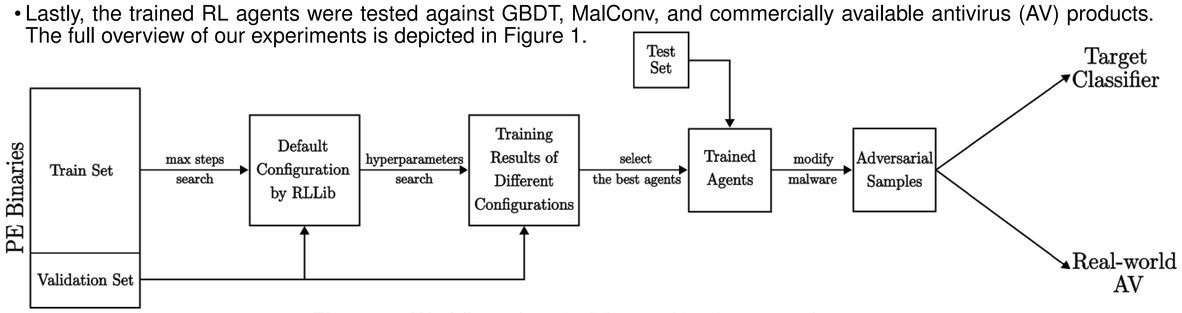

Figure 1: *Workflow of our training and testing procedure.*

4. Results

- The main metric used to evaluate adversarial attacks is the **evasion rate**, which represents the ratio of adversarial malware examples incorrectly classified as benign to the total number of files tested.

4.1 Results against ML-based Malware Classifiers

- We evaluated our best-trained agents (DQN, PG, and PPO) against the GBDT target classifier on a test set of 1,000 samples. The results are summarized in Table 1.
- **The highest evasion rate of 58.92% was achieved by the PPO agent while increasing the resulting file size by less than 10%.**

Table 1: *Results of the best configuration for each tested RL algorithm on the test set against the GBDT classifier. [%]*

	evasion rate	*size increase*
DQN	55.95	11.25
PG	40.14	**5.92**
PPO	**58.92**	9.01

Table 2: *Transferability of the adversarial attack targeted against GBDT to MalConv. [%]*

	GBDT	*MalConv*	*change*
DQN	55.95	27.17	-51.43
PG	40.14	18.96	-52.77
PPO	**58.92**	**28.91**	-50.93

- Subsequently, we tested the same three agents against the MalConv detector without any further training. The comparison of the results is shown in Table 2.
- **The PPO agent scored the highest evasion rate of 28.91%**, which represents a 50.93% decrease in comparison with the performance against the GBDT detector.

4.2 Results against Antivirus Products

- Finally, we tested the trained agents against nine best-rated AVs (e.g., Bitdefender, Avast, McAfee, ESET). The results are presented in Table 3.
- **The PPO agent achieved the highest average evasion rate of 13.85% against the AVs.**

Table 3: *Evasion rates of the generated adversarial samples against real-world AV programs. [%]*

	AV-1	*AV-2*	*AV-3*	*AV-4*	*AV-5*	*AV-6*	*AV-7*	*AV-8*	*AV-9*	*Average*
DQN	8.9	8.9	14.7	9.99	9.65	10.02	7.38	7.83	9.94	9.7
PG	9.25	9.25	15.0	9.35	9.27	9.92	16.86	12.5	9.6	11.22
PPO	**10.24**	**10.24**	**25.7**	**11.31**	**11.08**	**10.9**	**19.8**	**13.88**	**11.46**	**13.85**

5. Conclusion

- We proposed using reinforcement learning techniques to generate adversarial malware samples.
- We implemented ten binary file modifications that better preserve the original functionality of the malware in comparison with other RL-based generators.
- We optimized the hyperparameters of three RL algorithms (DQN, PG, and PPO) and tested them against two ML classifiers (GBDT and MalConv) and nine AVs.
- The best results were achieved by the PPO agent against the GBDT target classifier.
- When transferring the adversarial attack targeted against GBDT to MalConv and AVs, the evasion rate was significantly reduced but still not negligible as more than 10% of generated samples bypassed the AVs.
- **These results show that even commercially available AVs are vulnerable to adversarial attacks without the need for any knowledge of their internal structure.**
- A challenging future research area would be to design an adversarial attack capable of bypassing dynamic analysis methods.
- In addition, we plan to use the knowledge gained from the adversarial attacks to design new defense mechanisms against them.

References

[1] Anderson, H.S., Kharkar, A., Filar, B., Evans, D., Roth, P., 2018. Learning to evade static pe machine learning malware models via reinforcement learning. ArXiv abs/1801.08917. URL: http://adsabs.harvard.edu/abs/2018arXiv180108917A, doi:10.48550/arXiv.1801.08917.

[2] Song, W., Li, X., Afroz, S., Garg, D., Kuznetsov, D., Yin, H., 2022. Mab-malware: a reinforcement learning framework for blackbox generation of adversarial malware, in: Proceedings of the 2022 ACM on Asia Conference on Computer and Communications Security, Association for Computing Machinery, New York, NY, USA. pp. 990–1003. URL: https://doi.org/10.1145/3488932.3497768, doi:10.1145/3488932.3497768.