AI4SIM (AAG) Adversial Attack Generator

AAG is a modular tool designed to assess the resilience of AI models used in encrypted traffic analysis and intrusion detection by simulating adversarial scenarios. Its core functionality lies in injecting adversarial perturbations into clean network traffic features to evaluate model robustness under inference-time attacks. By supporting metrics such as Accuracy, F1-score, TPR, FPR, and an overall impact score, AAG enables a systematic analysis of model vulnerability and aids in identifying weaknesses before real-world deployment.

AI4CYBER

Component

The AAG solution provides:

- Improved AI Security Posture: AAG enables proactive testing of AI models against adversarial inputs, helping identify and mitigate vulnerabilities before deployment in real-world network environments.
- Modular and Extensible Architecture: With separate components for attack strategies, perturbation generation, and evaluation, AAG is easily extendable with new attack types, metrics, or model backends.
- Comprehensive Strategy Library: AAG includes a diverse set of white-box and black-box attack methods to simulate various adversarial scenarios, enabling thorough and comparative robustness evaluations.\
- User-Friendly Dashboard: AAG includes a visual, intuitive dashboard that allows users to configure attack parameters, run simulations, and view evaluation results interactively—making the tool accessible to both researchers and security analysts.\
- Actionable Insights for Risk Mitigation: The insights gained from AAG testing inform decision-makers and system architects about the adversarial resilience of models, supporting the secure integration of AI into critical infrastructure.

AAG (Adversarial Attack Generator) is a lightweight, modular tool designed to evaluate the robustness of AI models used in encrypted traffic analysis and intrusion detection. It leverages white box and black box adversarial attacks, measures performance degradation using key resilience metrics, and offers a user-friendly dashboard for easy configuration and analysis.

YouTube Video Link

The solution is still a prototype and the future tool will have commercial license.

Minds

Dimitris Asimopoulos (dasimopoulos@metamind.gr)
MetaMind Innovation (MINDS)

https://metamind.gr

- Asimopoulos, Dimitrios Christos, et al. “Breaching the Defense: Investigating FGSM and CTGAN Adversarial Attacks on IEC 60870-5-104 AI-Enabled Intrusion Detection Systems.” Proceedings of the 18th International Conference on Availability, Reliability and Security, ACM, 2023, pp. 1–8. DOI.org (Crossref),
- Asimopoulos, Dimitrios Christos, et al. “AAG: Adversarial Attack Generator for Evaluating the Robustness of Machine Learning Models against Adversarial Attacks.” 2024 IEEE International Conference on Big Data (BigData), IEEE, 2024, pp. 2682–89. DOI.org (Crossref)

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101070450. Disclaimer: Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the European Commission can be held responsible for them.