World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:19, No:01, 2025

Deep Reinforcement Learning and Generative Adversarial Networks Approach to Thwart Intrusions and Adversarial Attacks

Authors : Fabrice Setephin Atedjio, Jean-Pierre Lienou, Frederica F. Nelson, Sachin S. Shetty

Abstract : Malicious users exploit vulnerabilities in computer systems, significantly disrupting their performance and revealing the inadequacies of existing protective solutions. Even machine learning-based approaches, designed to ensure reliability, can be compromised by adversarial attacks that undermine their robustness. This paper addresses two critical aspects of enhancing model reliability. First, we focus on improving model performance and robustness against adversarial threats. To achieve this, we propose a strategy by harnessing deep reinforcement learning. Second, we introduce an approach leveraging generative adversarial networks to counter adversarial attacks effectively. Our results demonstrate substantial improvements over previous works in the literature, with classifiers exhibiting enhanced accuracy in classification tasks, even in the presence of adversarial perturbations. These findings underscore the efficacy of the proposed model in mitigating intrusions and adversarial attacks within the machine learning landscape.

Keywords : machine learning, reliability, adversarial attacks, deep-reinforcement learning, robustness
Conference Title : ICCSCIT 2025 : International Conference on Computer Science, Cybersecurity and Information Technology
Conference Location : Rome, Italy
Conference Dates : January 16-17, 2025