Research @ Botacin's Lab

Marcus Botacin[1]

[1]Texas A&M University (TAMU)
botacin@tamu.edu

Topics

Topics

Automated Attack Generation Using LLM models

GPThreats-3: Is Automatic Malware Generation a Threat?

Marcus Botacin
Texas A&M University
botacin@tamu.edu

***Abstract*—Recent research advances introduced large textual models, of which GPT-3 is state-of-the-art. They enable many applications, such as generating text and code. Whereas the model's capabilities might be explored for good, they might also cause some negative impact: The model's code generation capabilities might be used by attackers to assist in malware creation, a phenomenon that must be understood. In this work, our goal is to answer the question: Can current large textual models (represented by GPT-3) already be** attackers could use the models [10]. To contribute to this debate, we present an evaluation of the model's capabilities from the attacker's perspective. We explore how the models could assist attackers in many tasks, from the entire malware creation to the addition of anti-analysis techniques to existing code, and the automatic creation of malware variants via a scriptable procedure.

We investigated model capabilities by creating custom queries that were performed via OpenAI's public

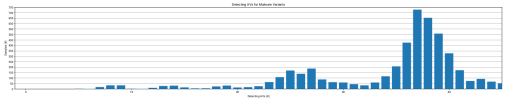


Figure 1: **Malware variants detection rates** vary according to the functions used to implement the same behaviors.

Source: `https://ieeexplore.ieee.org/document/10188649`

Topics

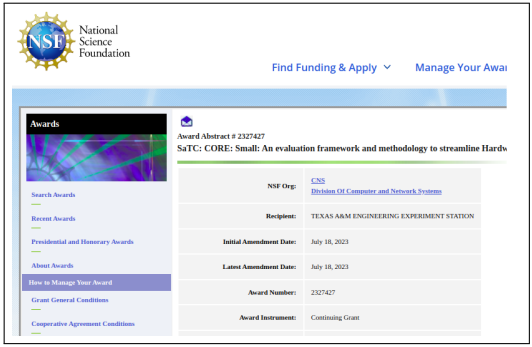
Adversarial ML in Practice

Figure: `mlsec.io`

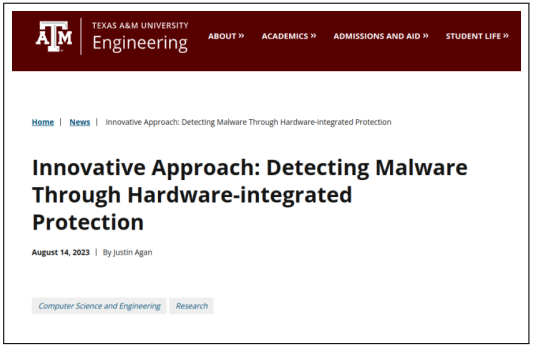
Figure: `https://cujo.com/machine-learning-security-evasion-competition-2020-results-and-behind-the-scenes/`

Topics

Hardware-Assisted Attack Detectors

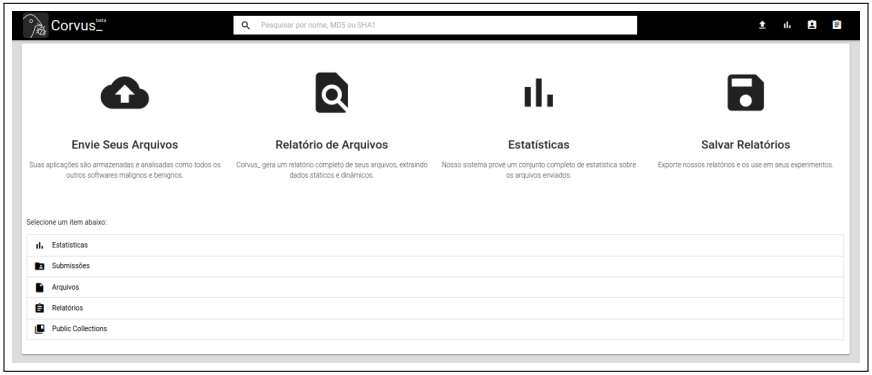


Source: `https://www.nsf.gov/awardsearch/showAward?AWD_ID=2327427`



Source: `tx.ag/ftOdCdj`

Threat Intelligence Platforms



Thanks!

botacin@tamu.edu