

Research @ Botacin's Lab

Marcus Botacin¹

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

Topics

- 1 Research Projects
 - Offensive Security
 - Offensive-Defensive Security
 - Defensive Security

Topics

- 1 Research Projects
 - Offensive Security

- Offensive-Defensive Security
- Defensive Security

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

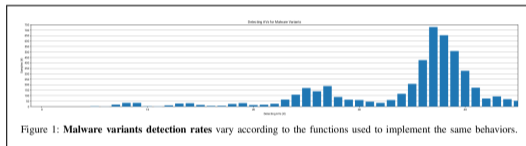


Figure 1: Malware variants detection rates vary according to the functions used to implement the same behaviors.

Source: <https://ieeexplore.ieee.org/document/10188649>

Topics

- 1 Research Projects
 - Offensive Security

- Offensive-Defensive Security
 - Defensive Security

Adversarial ML in Practice



Machine Learning Security Evasion Competition

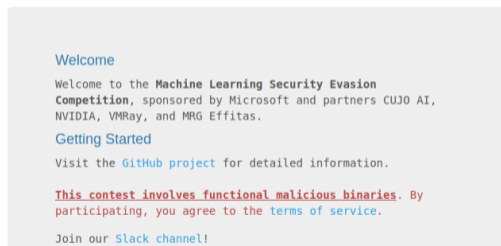


Figure: mlsec.io



Luckily, everyone understood this mistake and accepted the new results.

Analysis of the winning solutions

Please check out all the great write-ups from the participants.

First place in the attacker track and second at the defender track
<https://secret.inf.ufpr.br/2020/09/29/adversarial-malware-in-machine-learning-detectors-our-mlsec-2020-secrets/>

The previous one, but white-paper format, defender track only
<https://ieeexplore.ieee.org/document/8636415>

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

Topics

1 Research Projects

- Offensive Security

- Offensive-Defensive Security
- Defensive Security

Hardware-Assisted Attack Detectors



The screenshot shows the NSF Awards website interface. At the top left is the NSF logo and the text "National Science Foundation". To the right are navigation links: "Find Funding & Apply" and "Manage Your Award". Below this is a sidebar with a list of links: "Search Awards", "Recent Awards", "Presidential and Honorary Awards", "About Awards", "How to Manage Your Award" (highlighted), "Grant General Conditions", and "Cooperative Agreement Conditions". The main content area displays the award details for abstract # 2327427, titled "SaTC: CORE: Small: An evaluation framework and methodology to streamline Hardware-Assisted Attack Detectors".

NSF Org:	CNS Division Of Computer and Network Systems
Recipient:	TEXAS A&M ENGINEERING EXPERIMENT STATION
Initial Amendment Date:	July 18, 2023
Latest Amendment Date:	July 18, 2023
Award Number:	2327427
Award Instrument:	Continuing Grant

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



The screenshot shows the Texas A&M University Engineering website. The top navigation bar includes the TAMU logo, "TEXAS A&M UNIVERSITY Engineering", and links for "ABOUT", "ACADEMICS", "ADMISSIONS AND AID", and "STUDENT LIFE". Below the navigation bar is a breadcrumb trail: "Home | News | Innovative Approach: Detecting Malware Through Hardware-integrated Protection". The main heading is "Innovative Approach: Detecting Malware Through Hardware-integrated Protection". Below the heading is the date "August 14, 2023" and the author "By Justin Agan". At the bottom, there are tags for "Computer Science and Engineering" and "Research".

Source: tx.ag/ft0dCdj

Threat Intelligence Platforms

The screenshot displays the Corvus_ Threat Intelligence Platform interface. At the top, there is a navigation bar with the Corvus_ logo on the left and a search bar containing the text "Pesquisar por nome, MD5 ou SHA1". To the right of the search bar are icons for user profile, settings, and notifications. The main content area features four large, square tiles, each with an icon and a title:

- Envie Seus Arquivos**: Represented by a cloud with an upward arrow icon. Below the title, it states: "Suas aplicações são armazenadas e analisadas como todos os outros softwares malignos e benignos."
- Relatório de Arquivos**: Represented by a document with a magnifying glass icon. Below the title, it states: "Corvus_ gera um relatório completo de seus arquivos, extraindo dados estáticos e dinâmicos."
- Estatísticas**: Represented by a bar chart icon. Below the title, it states: "Nosso sistema provê um conjunto completo de estatística sobre os arquivos enviados."
- Salvar Relatórios**: Represented by a document with a circular icon. Below the title, it states: "Exporte nossos relatórios e os use em seus experimentos."

Below these tiles, there is a section titled "Selecione um item abaixo:" followed by a list of menu items, each with a small icon:

- Estadísticas
- Submissões
- Arquivos
- Relatórios
- Public Collections

Thanks!

botacin@tamu.edu