TechXplore

Using AI to develop enhanced cybersecurity measures

February 15 2024

Credit: Pixabay/CC0 Public Domain

A research team at Los Alamos National Laboratory is using artificial intelligence to address several critical shortcomings in large-scale malware analysis, making significant advancements in the classification of Microsoft Windows malware and paving the way for enhanced cybersecurity measures. Using their approach, the team has set a new world record in classifying malware families.

"Artificial intelligence methods developed for cyber-defense systems, including systems for large-scale malware analysis, need to consider real-world challenges," said Maksim Eren, a scientist in Advanced Research in Cyber Systems at Los Alamos. "Our method addresses several of them."

The team's paper was recently published in *ACM Transactions on Privacy and Security*.

This research introduces an innovative method using AI that is a significant breakthrough in the field of Windows malware classification. The approach achieves realistic malware family classification by leveraging semi-supervised tensor decomposition methods and selective classification, specifically, the reject option.

"The reject option is the model's ability to say 'I do not know,' instead of making a wrong decision, giving the model the knowledge discovery capability," Eren said.

Cyber defense teams need to quickly identify infected machines and malicious programs. These malicious programs can be uniquely crafted for their victims, which makes gathering large numbers of samples for traditional machine learning methods difficult.

This new method can accurately work with samples with both larger and smaller datasets at the same time—called class imbalance—allowing it to detect both rare and prominent malware families. It can also reject predictions if it is not confident in its answer. This could give security analysts the confidence to apply these techniques to practical high-stakes situations like cyber defense for detecting novel threats. Distinguishing between novel threats and known types of malware specimens is an essential capability to develop mitigation strategies. Additionally, this method can maintain its performance even when limited data is used in its training.

Altogether, the use of the reject option and tensor decomposition methods to extract multi-faceted hidden patterns in data, sets a superior capability in characterizing malware. This achievement underscores the groundbreaking nature of the team's approach.

"To the best of our knowledge, our paper sets a new world record by simultaneously classifying an unprecedented number of malware families, surpassing prior work by a factor of 29, in addition to operating under extremely difficult real-world conditions of limited data, extreme class-imbalance and with the presence of novel malware families," Eren said.

The team's tensor decomposition methods, with high performance computing and graphics processing unit capabilities, are now available as a user-friendly Python library in GitHub.

More information: Maksim E. Eren et al, Semi-Supervised Classification of Malware Families Under Extreme Class Imbalance via Hierarchical Non-Negative Matrix Factorization with Automatic Model Selection, *ACM Transactions on Privacy and Security* (2023). DOI: 10.1145/3624567

Provided by Los Alamos National Laboratory

Citation: Using AI to develop enhanced cybersecurity measures (2024, February 15) retrieved 10 May 2024 from https://techxplore.com/news/2024-02-ai-cybersecurity.html