1100 - Evading next-gen AV using artificial intelligence

Saturday at 11:00 in Track 4

20 minutes | Demo

Hyrum Anderson*Technical Director of Data Science, Endgame*

Much of next-gen AV relies on machine learning to generalize to never-before-seen malware. Less well appreciated, however, is that machine learning can be susceptible to attack by, ironically, other machine learning models. In this talk, we demonstrate an AI agent trained through reinforcement learning to modify malware to evade machine learning malware detection. Reinforcement learning has produced game-changing AI's that top human level performance in the game of Go and a myriad of hacked retro Atari games (e.g., Pong). In an analogous fashion, we demonstrate an AI agent that has learned through thousands of "games" against a next-gen AV malware detector which sequence of functionality-preserving changes to perform on a Windows PE malware file so that it bypasses the detector. No math or machine learning background is required; fundamental understanding of malware and Windows PE files is a welcome; and previous experience hacking Atari Pong is a plus.
Hyrum Anderson
Hyrum Anderson is technical director of data scientist at Endgame, where he leads research on detecting adversaries and their tools using machine learning. Prior to joining Endgame he conducted information security and situational awareness research as a researcher at FireEye, Mandiant, Sandia National Laboratories and MIT Lincoln Laboratory. He received his PhD in Electrical Engineering (signal and image processing + machine learning) from the University of Washington and BS/MS degrees from Brigham Young University. Research interests include adversarial machine learning, deep learning, large-scale malware classification, active learning, and early time-series classification.

#defcon25/by_track/track4/saturday #defcon25/By_Day/_saturday