1300 - A Picture is Worth a Thousand Words, Literally: Deep Neural Networks for Social Stego

Saturday at 13:00 in Track 4
45 minutes | Tool

Philip Tully*Principal Data Scientist, ZeroFOX*

Michael T. Raggo*Chief Security Officer, 802 Secure*

Images, videos and other digital media provide a convenient and expressive way to communicate through social networks. But such broadcastable and information-rich content provides ample illicit opportunity as well. Web-prevalent image files like JPEGs can be disguised with foreign data since they're perceivably robust to minor pixel and metadata alterations. Slipping a covert message into one of the billions of daily posted images may be possible, but to what extent can steganography be systematically automated and scaled?

To explore this, we first report the distorting side effects rendered upon images uploaded to popular social network servers, e.g. compression, resizing, format conversion, and metadata stripping. Then, we build a convolutional neural network that learns to reverse engineer these transformations by optimizing hidden data throughput capacity. From pre-uploaded and downloaded image files, the network learns to locate candidate metadata and pixels that are least modifiable during transit, allowing stored hidden payloads to be reliably recalled from newly presented images. Deep learning typically requires tons of training data to avoid over fitting. But data acquisition is trivial using social networks' free image hosting services, which feature bulk uploads and downloads of thousands of images at a time per album.

We show that hidden data can be predictably transmitted through social network images with high fidelity. Our results demonstrate that AI can hide data in plain sight, at large-scale, beyond human visual discernment, and despite third-party manipulation. Steganalysis and other defensive forensic countermeasures are notoriously difficult, and our exfiltration techniques highlight the growing threat posed by automated, AI-powered red teaming.
Philip Tully
Philip Tully is a Principal Data Scientist at ZeroFOX. He employs natural language processing and computer vision techniques in order to develop predictive models for combating security threats emanating from social networks. He earned his joint doctorate degree in computer science from the Royal Institute of Technology (KTH) and the University of Edinburgh, and has spoken at Black Hat, DEF CON , ShowMeCon and across the neuroscience conference circuit. He's a hackademic that's interested in applying brain-

inspired algorithms to both blue and red team operations.

@phtully

Michael T. Raggo

Michael T. Raggo, Chief Security Officer, 802 Secure (CISSP, NSA-IAM, CSI) has over 20 years of security research experience. His current focus is wireless IoT threats impacting the enterprise. Michael is the author of "Mobile Data Loss: Threats & Countermeasures" and "data Hiding: Exposing Concealed Data in Multimedia, Operating Systems, Mobile Devices and Network Protocols" for Syngress Books, and contributing author for "Information Security the Complete Reference 2nd Edition". A former security trainer, Michael has briefed international defense agencies including the FBI and Pentagon, is a participating member of FSISAC/BITS and PCI, and is a frequent presenter at security conferences, including Black Hat, DEF CON , Gartner, RSA, DoD Cyber Crime, OWASP, HackCon, and SANS.

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