Deep Learning for Real-time Malware Detection

Kate Highnam[1], Domenic Puzio[1]
[1]*Capital One*

Domain generation algorithm (DGA) malware makes callouts to unique web addresses to avoid detection by static rules engines. To counter this type of malware, we created an ensemble model that analyzes domains and evaluates if they were generated by a machine and thus potentially malicious. The ensemble consists of two deep learning models - a convolutional neural network and a long short-term memory network, both which were built using Keras and Tensorflow. These deep networks are flexible enough to learn complex patterns and do not require manual feature engineering. Deep learning models are also very difficult for malicious actors to reverse engineer, which makes them an ideal fit for cyber security use cases. The last piece of the ensemble is a natural-language processing model to assess whether the words in the domain make sense together. These three models are able to capture the structure and content of a domain, determining whether or not it comes from DGA malware with very high accuracy. These models have already been used to catch malware that vendor tools did not detect. Our system analyzes enterprise-scale network traffic in real time, renders predictions, and raises alerts for cyber security analysts to evaluate.