Training Malicious Domain Name Classifiers with Real, Heuristically Labeled Data

Bin Yu
CTO Office, Infoblox
Santa Clara, California
biny@infoblox.com

Daniel L. Gray
Institute of Technology
University of Washington, Tacoma
dangray@uw.edu

Martine De Cock*
Institute of Technology
University of Washington, Tacoma
mdecock@uw.edu

Anderson C. A. Nascimento
Institute of Technology
University of Washington, Tacoma
andclay@uw.edu

Abstract

Domain generation algorithms (DGAs) have become commonplace in malware that seek to establish command and control communication between an infected machine and the botmaster. DGAs dynamically generate large volumes of malicious domain names, only a few of which are registered by the botmaster and subsequently resolved when the malware on the infected machine tries to access them. Deep neural networks that can classify domain names as benign or malicious are of great interest in the fight against DGAs, because these networks can learn features themselves instead of having to rely on human engineered features. This is of particular importance in the cybersecurity domain, as malware constantly evolves. Keeping deep neural networks current for real-time detection of DGAs requires training them with massive up-to-date data. Since obtaining large amounts of clean ground truth labeled data is infeasible, an attractive alternative is the use of noise-not-free yet practical data collected based on the heuristic that most DGA domain names do not resolve. In this paper we compare a recurrent neural network trained on a small data set of ground truth data vs one trained on a data set of 50 million domain names obtained from real traffic via heuristic labeling.

*Guest professor at Ghent University, Dept. of Applied Mathematics, Computer Science and Statistics

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.