

A SYSTEM AND METHOD FOR DETECTING DOMAIN GENERATION ALGORITHMS (DGAs) USING DEEP LEARNING AND SIGNAL PROCESSING TECHNIQUES

Abstract

This document describes a system and method for detecting domain names that exhibit Domain Generation Algorithm (DGA) like behaviours from a stream of Domain Name System (DNS) records. In particular, this document describes a system comprising a deep learning classifier (DL-C) module for receiving and filtering the stream of DNS records before the filtered DNS records, which have been determined to possess domain names that exhibit DGA behaviour are provided to a series filter-classifier (SFC) module. The SFC module then groups the records into various series based on source IP, destination IP and time. For each series, it then filters away records that do not exhibit the dominant DGA characteristics of the series. Finally, for each series, it makes use of the remaining DNS records' timestamps to generate a time series of DGA occurrences and then, using this time series of occurrences, determine the number of DGA bursts throughout the time period of analysis. An autoencoder-classifier (AE-C) then assigns coherence scores, by analysing correlations over a time period, to each series of DGA records based on its corresponding time series of DGA occurrences. A frequency spectrum analyser (FSA) module is then used to convert the time series of DGA occurrences into a frequency spectrum, before it identifies periodic DGA bursts occurring within each series of DGA records. The information generated by the FSA and AE-C modules together with the series of possible DGA domain names produced by the SFC module and other enriching details are then passed to an alert module which then uses this information to present and prioritize enriched DGA alerts.

FIGURE 1