BOTNET - DEEP LEARNING BASED APPROACHES

DATE : 11th May’ 2023 **BY : NEEL SHAH**

- I have analyzed content available on the internet about botnet detection and I found that for botnet detection there are various algorithms types such as supervised learning, unsupervised learning and semi-supervised learning but in this study I have focused on the deep learning approach.
- But while studying / researching about this topic i have also found and i want to explicitly mention that the supervised learning approach specifically tree and ensemble based algorithm sawed quite good performance or nearly same performance to the deep learning approaches, And one of the advantage of tree / ensemble based algorithm requires less computational power as compared to deep learning algorithms.
- One disadvantage of using a deep learning based approach is that we would require more computational power.
- Basically top deep learning approaches or algorithms are:
    - Convolutional Neural Networks (CNNs)
    - Long Short Term Memory Networks (LSTMs)
    - Recurrent Neural Networks (RNNs)
    - Generative Adversarial Networks (GANs)
    - Radial Basis Function Networks (RBFNs)
    - Multilayer Perceptrons (MLPs)
    - Self Organizing Maps (SOMs)
    - Deep Belief Networks (DBNs)
    - Restricted Boltzmann Machines(RBMs)
- Overall i found around 7 approaches which are based on the deep learning and details about them are as mentioned below:

Sr. No.	Name Of Paper	Result	Approach Used
1.	DeepBot: a time-based botnet detection with deep learning	99.36%	RNN and LSTM
2.	Automating Botnet detection with graph neural networks	99.5%	GNN
3.	DBD : Deep Learning DGA-Based Botnet Detection	97.80%	DGA(Domain Generation Algorithm)

			contains CNN, LSTM
4.	BotShark: A Deep Learning Approach for Botnet Traffic Detection	TPR: 0.91 FPR: 0.13	Autoencoder and CNN
5.	VARMAN: Multi-plane security framework for software defined networks	96%	Stacked auto encoders and random forest
6.	Machine Learning-Based IoT-Botnet Attack Detection with Sequential Architecture	99%	ANN, J48 decision tree, Naive Bayes
7.	Mobile Botnet Detection: A Deep Learning Approach Using Convolutional Neural Networks	98.9%	CNN

- An trend is also seen that hybrid approaches i.e model containing more than one algorithms show higher performance as compared to the single deep learning models
- Brief details about the above mentioned algorithms are as written below in the sequential order as that in table:
  1. They chose to apply RNN and LSTM on time-based analysis to detect potential botnets in this paper.
     → In the first step they extract features from the input network traffic and then, in step two, they train using RNN and LSTM.
     → They also used PyShark in their approach which is an important tool for network packet analysis.
  2. In this work, they proposed to tailor graph neural networks (GNN) to identify botnets within massive background Internet communication graphs by automatically identifying their topological features (i.e., communication patterns).
     → They specifically design GNNs for the problem of botnet detection that can automatically capture the hierarchical structure of centralized botnets and the fast-mixing structure for decentralized botnets.
     → They formulate this problem as a binary node classification problem on graphs.
  3. They proposed a scalable deep learning DGA-based botnet identification framework, named, DBD, which uses a Convolutional Neural Network with a Long Short Term Memory (CNN-LSTM) pipeline.
  4. They proposed BoTShark with two deep structures namely BoTShark-SA that applies stacked Autoencoders to extract new features for distinguishing malicious and benign network flows and BoTShark-CNN which takes the advantage of CNNs to train a classifier for detecting malicious traffic.

→ An Autoencoder is similar to a neural network where the output is regarded as the input and supposes that the hidden layer must reconstruct the initial information with the least possible amount of distortion. Then, Autoencoders are trained to reconstruct their own inputs X instead of being trained to predict Y . An Autoencoder tries to learn a function h(W,b) (x) ~= x by learning an approximation to the identity function.
→ Also in the autoencoders they used the softmax activation function.

5. They used stacked auto encoders and random forest classifiers.
   → It integrates and coordinates multiple stages of attack detection including feature selection and flow classification, anomaly detection to detect novel intrusions with a self-learning ability.
6. This approach uses an artificial neural network (ANN), J48 decision tree, and Naïve Bayes.
   → This is a hybrid approach.
7. It is purely CNN based approach
   → They performed an experiment using the different numbers of the layers and they finally found that for their usage when the number of layers were equal to 32 it gives maximum accuracy.

- At the end I suggest using hybrid model based approaches as they give the highest accuracy.