University of Ottawa
Faculty of Engineering
School of Electrical Engineering and Computer Science

Assignment 1

Name: Gehad Hisham Hassan Abdelghany Somaa.
Student ID: 300327296

The authors of the paper:
Encrypt DNS Traffic: Automated Feature Learning Method for Detecting DNS Tunnels.

The title of the paper:
Shuai Ding, Daoqing Zhang, Jingguo Ge, Xiaowei Yuan, Xinhui Du.

Summary

The paper tries to solve that the DNS can be manipulated as the attackers use the DNS protocol itself to collect, exfiltrate and transport personal information through DNS tunnel. The authors developed a deep learning model to classify normal DNS and anomaly DNS tunnels traffic. The model relies on an attention mechanism in addition to a bidirectional GRU-based network. The authors run their tests on public data set that contains both normal and abnormal DNS traffic. This was using traditional machine learning and deep learning models. The outcomes of the experiments show that the model developed by the authors has a better performance than machine learning and other deep learning models with detecting the DNS tunnels.

Critical review

Research Goal

The goal of the research is to detect the malicious DNS tunnels. DNS tunnel gives the means of encapsulating data within DNS requests and take advantage of the free flow of DNS traffic. The traditional DNS tunnel detection is no longer useful because of various DNS encryption methods used by attackers. The authors of the paper develop a new deep learning method to detect anomalies in the data.

Clarity

The paper written in a way that needs knowledge in deep learning, related to attention mechanism, not only machine learning, besides a strong knowledge with cybersecurity, in a way that requires to search outside the paper itself to understand clearly what the problem the paper discusses and how solves it.

Related Works

The authors make a good job in the related work section, they discuss it from different parts, provide several related papers and works from different perspectives for the same problem and provides these works in the references section.

Methods

The authors developed a deep learning model [ABG-VAE] to classify normal DNS and anomaly DNS tunnels traffic that uses a variational auto encoder which depends on attention mechanism to handle row flow sequential data, in addition to a bidirectional GRU-based VAE model to handle long sequential and structural data, and it is not dependent on packet content like traditional methods. and uses packet sequence features consists of three dimensions which have the most information about the traffic behavior. They use two different data set, one contains normal traffic data to enable the model to detect unknown types of DNS tunnels, and the other dataset contains No-DoH traffic, normal DoH traffic and malicious traffic.

Results and Claims

The authors model has better performance and is less complicated than other models provided in the paper comparison, they compare their model with other machine learning and deep learning models, and compare their accuracies, recall, precision, and F1 scores, also their model has fewer parameters and shorter training. The authors use some features that can affect the result of the model and found out that the sequence length has an impact on the model performance. They made another experiment to ensure the important of the attention mechanism, and find out that the threshold between normal and malicious data can be more obvious with attention.

Support of Results and Claims

The authors compare their deep learning model [ABG-VAE] with other machine learning models, RF is the only model that has better performance, but it is more complicated and time consuming, as it requires feature engineering and features manual extracting, and on the other hand, their model does not need features engineering, as from the raw flow sequences, it learns the representative features. And they also compared their model with deep learning models, which was not able to detect unknown abnormal DNS traffic tunnels, only their model was able to detect those traffics.

Missing Claims and Results

This work only uses training and testing dataset, and don't consider using validation dataset, which is very useful while tuning model hyperparameters, and it would be better to say how can they do the future work not just mention it.

Discussion

The paper discussed DNS tunnels detection from different fields, but it requires you to understand some concepts before reading the paper which can be a limitation in the discussion of the paper. But overall, the discussion is well written as it discusses the problem from different perspectives, which consider a very strong point for the paper.

Future Work

The paper says that the attackers may use other techniques to encode the malicious data like packet padding, that make it hard to detect, but didn't consider in their paper how will this happen.
The future work that can be considered is to use deep learning models, like CNN, which has a good performance in this field, beside transformers models, and pay attention to other features as it can affect the detection performance.