Utilizing Big Data Analytics for Cybersecurity

Joshua Joseph

Introduction

- Number of cyber attacks on system networks have dramatically increased over the past several years
- There is a growing need for additional tools to better mitigate and understand security threats
- Both consumers and businesses could benefit from a better understanding of the cyber threats they face

Topic Domain

- The project falls under the domain of network security
- More specifically within network security, the paper evaluates IDS and the type of traffic that goes through the network
- Cross sectionality between network security and big data tools such as data analytics and machine learning

Dataset

- The dataset simulates a military's network
- Simulated US Air Force LAN with various attacks
- To create the dataset an environment that acquired raw TCP/IP packets were used
- Consists of Training and Testing sets
- Date created 2018-10-09
- Derived from Kaggle
- Test set contains
  - 22,544 rows & 41 columns
- Training set contains
  - 25,192 rows & 42 columns
- Link: https://www.kaggle.com/sampadab17/network-intrusion-detection

Solution

- Big data tools allows for better understanding of threats because of the ability to visualize, model, and efficiently summarize the data
- Create ML prediction model based on dataset to predict threats
- Understand common attributes for threats
- Compare efficiency of various models in predicting threats