CIC-IDS-2017 dataset

CIC-IDS-2017 dataset is created by Canadian Institute for Cybersecurity at the University of New Brunswick. The authors of dataset argue that most of the dataset since 1998 are out of date and unreliable. Some of these datasets suffer from the lack of traffic diversity and volumes, some do not cover the variety of known attacks, while others anonymize packet payload data, which cannot reflect the current trends. Some are also lacking feature set and metadata.

The dataset contains benign and the most up-to-date common attacks, which resembles the true real-world data (PCAPs). It also include the results of the network traffic analysis using CICFlowMeter with labeled flows based on the timestamp, source IP, destination IP, source and destination ports, protocol and attack. Generating realistic background traffic was their top priority. They use their proposed B-profile system to profile the abstract behavior of human interactions and generates naturalistic benign background traffic. They build the abstract behaviour of 25 users based on HTTP, HTTPS, FTP, SSH and emails protocol.

The data capturing period started at 9 a.m Monday, July 3, 2017 and ended at 5p.m Friday, July, 2017. Monday is the normal day and only contain benign traffic. The implemented attack include Brute Force FTP, Brute Force SSH, Dos, Heartbleed, Web Attack, Infiltration, Botnet and DDoS. They have been executed both morning and afternoon on Tuesday, Wednesday, Thursday and Friday. The dataset cover 11 criteria which are proposed for IDS benchmark dataset, none of the previous dataset cover all. The criteria include: complete network configuration, complete traffic, labelled dataset, complete interaction, complete capture, available protocols, attack diversity, heterogeneity, feature set and metadata.