CIC-IDS-2017 dataset

The Canadian Institute for Cybersecurity recognizes the significant limitations of many networks intrusion detection (IDS) dataset since 1998. These limitations often include outdated attacks models, limited traffic diversity, incomplete attack coverage and lack of feature sets and metadata. To address these shortcomings, CIC has developed the CIC-IDS-2017 dataset. This comprehensive dataset offers several key advantages: Real-World Reflectance, Network Insights, Naturalistic Background Traffic. The network traffic was captured over a five-day period, starting at 9 a.m. on Monday, July 3, 2017, and concluding at 5 p.m. on Friday, July 7, 2017. Monday’s capture exclusively comprised benign traffic, while a spectrum of attacks was executed during working hours on Tuesday, Wednesday, Thursday, and Friday. These attacks encompassed Brute Force FTP, Brute Force SSH, Dos, Heartbleed, Web Attack, Infiltration, Botnet and DDoS.

The CIC-IDS-2017 dataset meticulously analyzes network flows using CICFlowMeter, yielding a comprehensive set of 84 features that provide a granular overview of network activity. These features can be broadly categorized into:

Fundamental Flow Information: This includes essential details such as IP addresses, timestamps, protocols used, and the duration of each flow.

Packet-Level Insights: These features delve into the number of packets exchanged and their size characteristics, offering insights into the granularity of network communication.

Flow-Specific Statistics: This category encompasses both byte-level statistics, revealing data volume and directionality, as well as time-related statistics, shedding light on flow duration and idle periods.

Inter-Arrival Time Patterns: These features calculate the time gaps between consecutive packets within a flow, potentially unmasking unusual patterns associated with malicious activities.

Flag Distribution: The frequency of specific flags (SYN, FIN, RST, PSH, ACK) within a flow provides valuable insights into the communication protocol and potential deviations from standard patterns.

For a comprehensive exploration of each feature, please refer to reference [3].