Network Intrusion Detection DATASET

CICIDS2017 (Canadian Institute for Cybersecurity Intrusion Detection System 2017) :[1]

1.The CICIDS2017 dataset is a labelled dataset, meaning that each instance in the dataset has been labelled as either benign or malicious traffic. The dataset includes nine different types of attacks, and each attack is labelled with a unique identifier. In addition to the attack labels, the dataset also includes labels for the protocol used in each instance, as well as labels for the source and destination IP addresses and ports.

2.The labelling of the CICIDS2017 dataset was performed by a team of cybersecurity experts at the Canadian Institute for Cybersecurity (CIC) at the University of New Brunswick. The labelling process involved analysing each instance in the dataset to determine whether it was benign or malicious traffic, and if it was malicious, to identify the specific type of attack.

3.The dataset contains approximately 2.8 million network traffic instances and includes nine different types of attacks, including DoS, DDoS, port scanning, and data exfiltration. The dataset also includes a variety of benign traffic, including HTTP, FTP, FTP, SSH, email protocols and DNS traffic.

4. The data capturing period started at 9 a.m., Monday, July 3, 2017, and ended at 5 p.m. on Friday, July 7, 2017, for a total of 5 days. Monday is the normal day and only includes benign traffic. The implemented attacks 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.

5. The dataset is provided in PCAP format, which is a common file format for capturing network traffic

UNSW-NB15 (University of New South Wales Network-Based 15) : [2]

1. is a dataset that contains network traffic data captured in a simulated network environment. It includes both benign and malicious traffic, and the malicious traffic includes a variety of different attack types
2. The total number of records is two million and 540,044 which are storedin the four CSV files, namely, UNSW-NB15_1.csv, UNSW-NB15_2.csv, UNSW-NB15_3.csv and UNSW-NB15_4.csv.

CICIDS2017 vs UNSW-NB15:

- They differ in terms of the environment in which the data was captured and the types of attacks that are included. UNSW-NB15 was captured in a simulated environment, while CICIDS2017 was captured in a real-world environment. UNSW-NB15 includes more attack types, while CICIDS2017 includes a larger number of instances.

Some other network intrusion detection datasets that are commonly used in research: [3,4,5,6]

1. KDD Cup 1999 dataset: This is one of the earliest and most widely used datasets for network intrusion detection. It contains network traffic data from a simulated environment and includes both benign and malicious traffic.
2. UNSW-NB 10 dataset: This is a smaller version of the UNSW-NB15 dataset that includes the same types of attacks but with fewer instances.
3. DARPA Intrusion Detection Evaluation dataset: This dataset was created by the Defence Advanced Research Projects Agency (DARPA) to evaluate intrusion detection systems. It includes network traffic data from a simulated environment and is labelled with information about the type of attack.

4. ISCX IDS 2012 dataset: This dataset was created by the Information Security Centre of Excellence (ISCX) to evaluate intrusion detection systems. It includes network traffic data from a real-world environment and is labelled with information about the type of attack.

References:

1.CICIDS2017 [https://www.unb.ca/cic/datasets/ids-2017.html]

2. UNSW_15 [https://research.unsw.edu.au/projects/unsw-nb15-dataset]

3. https://www.unb.ca/cic/datasets/nsl.html

4. KDD CUP 99 [https://archive.ics.uci.edu/ml/datasets/kdd+cup+1999+data]

5. DARPA dataset [https://www.ll.mit.edu/r-d/datasets/1998-darpa-intrusion-detection-evaluation-dataset]

6. ISCX-IDS dataset [http://www.iscx.ca/dataset]