

In [44]: `import pandas as pd`

In [45]: `url = "data/Friday-WorkingHours-Afternoon-DDos.pcap_ISCX.csv"
da = pd.read_csv(url)`

In [102]: `da`

Out[102]:

| | Flow ID | Source IP | Source Port | Destination IP | Destination Port | Protocol | Timestamp | Flow Duration | Total Fwd Packets | Total Backward Packets | ... | Active Mean | |
|--------|---------|---|----------------|----------------|------------------|----------|-----------|---------------|-------------------|------------------------|-----|-------------|-----|
| | 0 | 192.168.10.5-104.16.207.165-54865-443-6 | 104.16.207.165 | 443 | 192.168.10.5 | 54865 | 6 | 7/7/2017 3:30 | 3 | 2 | 0 | ... | 0.0 |
| | 1 | 192.168.10.5-104.16.28.216-55054-80-6 | 104.16.28.216 | 80 | 192.168.10.5 | 55054 | 6 | 7/7/2017 3:30 | 109 | 1 | 1 | ... | 0.0 |
| | 2 | 192.168.10.5-104.16.28.216-55055-80-6 | 104.16.28.216 | 80 | 192.168.10.5 | 55055 | 6 | 7/7/2017 3:30 | 52 | 1 | 1 | ... | 0.0 |
| | 3 | 192.168.10.16-104.17.241.25-46236-443-6 | 104.17.241.25 | 443 | 192.168.10.16 | 46236 | 6 | 7/7/2017 3:30 | 34 | 1 | 1 | ... | 0.0 |
| | 4 | 192.168.10.5-104.19.196.102-54863-443-6 | 104.19.196.102 | 443 | 192.168.10.5 | 54863 | 6 | 7/7/2017 3:30 | 3 | 2 | 0 | ... | 0.0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 225740 | | 192.168.10.15-72.21.91.29-61374-80-6 | 72.21.91.29 | 80 | 192.168.10.15 | 61374 | 6 | 7/7/2017 5:02 | 61 | 1 | 1 | ... | 0.0 |
| 225741 | | 192.168.10.15-72.21.91.29-61378-80-6 | 72.21.91.29 | 80 | 192.168.10.15 | 61378 | 6 | 7/7/2017 5:02 | 72 | 1 | 1 | ... | 0.0 |
| 225742 | | 192.168.10.15-72.21.91.29-61375-80-6 | 72.21.91.29 | 80 | 192.168.10.15 | 61375 | 6 | 7/7/2017 5:02 | 75 | 1 | 1 | ... | 0.0 |
| 225743 | | 192.168.10.15-8.41.222.187-61323-80-6 | 8.41.222.187 | 80 | 192.168.10.15 | 61323 | 6 | 7/7/2017 5:02 | 48 | 2 | 0 | ... | 0.0 |
| 225744 | | 192.168.10.15-8.43.72.21-61326-80-6 | 8.43.72.21 | 80 | 192.168.10.15 | 61326 | 6 | 7/7/2017 5:02 | 68 | 1 | 1 | ... | 0.0 |

225745 rows × 86 columns

In [56]: `da.head()`

Out[56]:

| | Flow ID | Source IP | Source Port | Destination IP | Destination Port | Protocol | Timestamp | Flow Duration | Total Fwd Packets | Total Backward Packets | ... | min_seg_si |
|---|---|----------------|-------------|----------------|------------------|----------|---------------|---------------|-------------------|------------------------|-----|------------|
| 0 | 192.168.10.5-104.16.207.165-54865-443-6 | 104.16.207.165 | 443 | 192.168.10.5 | 54865 | 6 | 7/7/2017 3:30 | 3 | 2 | 0 | ... | |
| 1 | 192.168.10.5-104.16.28.216-55054-80-6 | 104.16.28.216 | 80 | 192.168.10.5 | 55054 | 6 | 7/7/2017 3:30 | 109 | 1 | 1 | ... | |
| 2 | 192.168.10.5-104.16.28.216-55055-80-6 | 104.16.28.216 | 80 | 192.168.10.5 | 55055 | 6 | 7/7/2017 3:30 | 52 | 1 | 1 | ... | |
| 3 | 192.168.10.16-104.17.241.25-46236-443-6 | 104.17.241.25 | 443 | 192.168.10.16 | 46236 | 6 | 7/7/2017 3:30 | 34 | 1 | 1 | ... | |
| 4 | 192.168.10.5-104.19.196.102-54863-443-6 | 104.19.196.102 | 443 | 192.168.10.5 | 54863 | 6 | 7/7/2017 3:30 | 3 | 2 | 0 | ... | |

5 rows × 85 columns

In [103]: `da.shape`

Out[103]: (225745, 86)

In [104]: `da.columns`

Out[104]: Index(['Flow ID', 'Source IP', 'Source Port', 'Destination IP', 'Destination Port', 'Protocol', 'Timestamp', 'Flow Duration', 'Total Fwd Packets', 'Total Backward Packets', 'Total Length of Fwd Packets', 'Total Length of Bwd Packets', 'Fwd Packet Length Max', 'Fwd Packet Length Min', 'Fwd Packet Length Mean', 'Fwd Packet Length Std', 'Bwd Packet Length Max', 'Bwd Packet Length Min', 'Bwd Packet Length Mean', 'Bwd Packet Length Std', 'Flow Bytes/s', 'Flow Packets/s', 'Flow IAT Mean', 'Flow IAT Std', 'Flow IAT Max', 'Flow IAT Min', 'Fwd IAT Total', 'Fwd IAT Mean', 'Fwd IAT Std', 'Fwd IAT Max', 'Fwd IAT Min', 'Bwd IAT Total', 'Bwd IAT Mean', 'Bwd IAT Std', 'Bwd IAT Max', 'Bwd IAT Min', 'Fwd PSH Flags', 'Bwd PSH Flags', 'Fwd URG Flags', 'Bwd URG Flags', 'Fwd Header Length', 'Bwd Header Length', 'Fwd Packets/s', 'Bwd Packets/s', 'Min Packet Length', 'Max Packet Length', 'Packet Length Mean', 'Packet Length Std', 'Packet Length Variance', 'FIN Flag Count', 'SYN Flag Count', 'RST Flag Count', 'PSH Flag Count', 'ACK Flag Count', 'URG Flag Count', 'CWE Flag Count', 'ECE Flag Count', 'Down/Up Ratio', 'Average Packet Size', 'Avg Fwd Segment Size', 'Avg Bwd Segment Size', 'Fwd Header Length.1', 'Fwd Avg Bytes/Bulk', 'Fwd Avg Packets/Bulk', 'Fwd Avg Bulk Rate', 'Bwd Avg Bytes/Bulk', 'Bwd Avg Packets/Bulk', 'Bwd Avg Bulk Rate', 'Subflow Fwd Packets', 'Subflow Fwd Bytes', 'Subflow Bwd Packets', 'Subflow Bwd Bytes', 'Init_Win_bytes_forward', 'Init_Win_bytes_backward', 'act_data_pkt_fwd', 'min_seg_size_forward', 'Active Mean', 'Active Std', 'Active Max', 'Active Min', 'Idle Mean', 'Idle Std', 'Idle Max', 'Idle Min', 'Label', 'label'], dtype='object')

In [101]: `# da["Source Port"] # error`

In [106]: `# da.loc[:, ["Source IP", "Source Port", "Destination IP", "Destination Port", "Protocol", "Label"]]`

In [91]: `da.iloc[:, [1,2,3,4,5,84]] # displaying columns by iloc`

Out[91]:

| | Source IP | Source Port | Destination IP | Destination Port | Protocol | Label |
|--------|----------------|-------------|----------------|------------------|----------|--------|
| 0 | 104.16.207.165 | 443 | 192.168.10.5 | 54865 | 6 | BENIGN |
| 1 | 104.16.28.216 | 80 | 192.168.10.5 | 55054 | 6 | BENIGN |
| 2 | 104.16.28.216 | 80 | 192.168.10.5 | 55055 | 6 | BENIGN |
| 3 | 104.17.241.25 | 443 | 192.168.10.16 | 46236 | 6 | BENIGN |
| 4 | 104.19.196.102 | 443 | 192.168.10.5 | 54863 | 6 | BENIGN |
| ... | ... | ... | ... | ... | ... | ... |
| 225740 | 72.21.91.29 | 80 | 192.168.10.15 | 61374 | 6 | BENIGN |
| 225741 | 72.21.91.29 | 80 | 192.168.10.15 | 61378 | 6 | BENIGN |
| 225742 | 72.21.91.29 | 80 | 192.168.10.15 | 61375 | 6 | BENIGN |
| 225743 | 8.41.222.187 | 80 | 192.168.10.15 | 61323 | 6 | BENIGN |
| 225744 | 8.43.72.21 | 80 | 192.168.10.15 | 61326 | 6 | BENIGN |

225745 rows × 6 columns

In [98]: `da.iloc[:, [1,2,3,4,5,84]].values # five-tuple labelled`

Out[98]: array([['104.16.207.165', 443, '192.168.10.5', 54865, 6, 'BENIGN'],
['104.16.28.216', 80, '192.168.10.5', 55054, 6, 'BENIGN'],
['104.16.28.216', 80, '192.168.10.5', 55055, 6, 'BENIGN'],
...,
['72.21.91.29', 80, '192.168.10.15', 61375, 6, 'BENIGN'],
['8.41.222.187', 80, '192.168.10.15', 61323, 6, 'BENIGN'],
['8.43.72.21', 80, '192.168.10.15', 61326, 6, 'BENIGN']],
dtype=object)

In [107]: `fa = da.iloc[:, [1,2,3,4,5,84]].values
fa[0:5]`

Out[107]: array([['104.16.207.165', 443, '192.168.10.5', 54865, 6, 'BENIGN'],
['104.16.28.216', 80, '192.168.10.5', 55054, 6, 'BENIGN'],
['104.16.28.216', 80, '192.168.10.5', 55055, 6, 'BENIGN'],
['104.17.241.25', 443, '192.168.10.16', 46236, 6, 'BENIGN'],
['104.19.196.102', 443, '192.168.10.5', 54863, 6, 'BENIGN']],
dtype=object)

In []: