**Dataset overview**

With the expanded applications of modern-day networking, network infrastructures are at risk from cyber-attacks and intrusions. Multiple datasets have been proposed in the literature that can be used to create Machine Learning (ML) based Network Intrusion Detection Systems (NIDS). However, many of these datasets suffer from sub-optimal performance and do not adequately and effectively represent all types of intrusions. Another problem with these datasets is the low accuracy of tail classes. To address these issues, the University of Nevada - Reno proposed an Intrusion Detection Dataset (UNR-IDD) that provides researchers with a wider range of samples and scenarios.

**Features**

UNR-IDD dataset has 21 features as below:

* Switch ID: The switch through which the network flow passed
* Port Number: The switch port through which the flow passed
* Received packets: Number of packets received by the port
* Received bytes: Number of bytes received by the port
* Sent bytes: Number of bytes sent by the port
* Sent packets: Number of packets sent by the port
* Port alive duration (s) : The time port has been alive in seconds
* Delta received packets: The delta port statistics are computed on the controller side by taking the difference between the last two collected data instances.
* Delta received bytes
* Delta sent bytes
* Delta received packets
* Delta port alive duration (s)
* Connection Point: Network connection point expressed as a pair of the network element identifier and port number.
* Total Load/Rate: Obtain the current observed total load/rate (in bytes/s) on a link.
* Total Load/Latest: Obtain the latest total load bytes counter viewed on that link.
* Unknown Load/Rate: Obtain the current observed unknown-sized load/rate (in bytes/s) on a link.
* Unknown Load/Latest: Obtain the latest unknown-sized load bytes counter viewed on that link.
* Latest bytes counter: Latest bytes counted in the switch port
* Active Flow Entries: Returns the number of active flow entries in this table.
* Packets Looked Up: Returns the number of packets looked up in the table.
* Packets Matched: Returns the number of packets that successfully matched in the table.

**Labels**

The following intrusions are simulated in the data collection phase:

* TCP-SYN Flood: A Distributed Denial of Service (DDoS) attack where attackers target hosts by initiating many Transmission Control Protocol (TCP) handshake processes without waiting for the response from the target node. By doing so, the target device's resources are consumed as it has to keep allocating some memory space for every new TCP request.
* Port scan: An attack in which attackers scan available ports on a host device to learn information about the services, versions, and even security mechanisms that are running on that host.
* Flow Table Overflow: An attack that targets network switches/routers where attacks compromise the functionality of a switch/router by consuming the flow tables that forward packets with illegitimate flow entries and rules so that legitimate flow entries and rules cannot be installed.
* Blackhole: An attack that targets network switches/routers to discard the packets that pass through, instead of relaying them on to the next hop.
* Traffic Diversion: An attack that targets network switches/routers to reroute the direction of packets away from their destination, intending to increase travel time and/or spying on network traffic through a man-in-the-middle scenario.
* Normal Traffic

These intrusion types were selected for this dataset as they are common cyber attacks that can occur in any networking environment. Also, these intrusion types cover attacks that can be launched on both network devices and end hosts. This dataset can be defined as a multi-class classification problem.

The provided dataset is divide into three different subsets with the following percentages:

* 70% for training
* 15% for validation
* 15% for testing