Modern computer networks and their connected applications have reached unprecedented growth with implementations like the internet of things, smart homes, and software-defined networks. However, this has also increased the potential for network intrusions, which are a continuous threat to network infrastructures as they attempt to compromise the major principles of computing systems: availability, authority, confidentiality, and integrity. These threats are difficult to detect unaided, as they display indistinguishable network traffic patterns as normal functionality. To provide enhanced protection against intrusions, the usage of machine learning for NIDS has gained traction in the last decade as various open-sourced datasets have been proposed and established by multiple research groups globally.

However, a common problem that has been identified with many of these datasets is inadequate modeling of tail classes. Another limitation of the current datasets is that they mostly depend on flow level statistics, which can limit the transferability of the NIDS solutions to other network configurations. Lastly, some existing datasets suffer from incomplete or missing records. These records or samples must be ignored or dropped from the overall dataset, which leads to sub-optimal performance.

The main difference between UNR-IDD and existing datasets is that UNR-IDD consists primarily of network port statistics. These refer to the observed port metrics recorded in switch/router ports within a networking environment. The dataset also includes delta port statistics which indicates the change in magnitude of observed port statistics within a time interval. Compared to datasets that primarily use flow level statistics, these port statistics can provide a fine-grained analysis of network flows from the port level as decisions are made at the port level versus the flow level. This can lead to rapid identification of potential intrusions. We also address the limitation of the presence of tail classes. Our dataset ensures that there are enough samples for ML classifiers to achieve high F-Measure scores, uniquely. Our proposed dataset also ensures that there are no missing network metrics and that all data samples are filled.

This dataset contains 5000 records of features extracted from Network Port Statistics to protect modern-day computer networks from cyber attacks and are thereby classified into 5 classes

Normal:0,

Blackhole:1,

TCP-SYN:2,

PortScan:3,

Diversion:4,

Overflow:5

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

Packets Rx Dropped - Number of packets dropped by the receiver

Packets Tx Dropped - Number of packets dropped by the sender

Packets Rx Errors - Number of transmit errors

Delta Received Packets - Number of packets received by the port

Delta Received Bytes - Number of bytes received by the port

Delta Sent Bytes - Number of bytes sent by the port

Delta Sent Packets - Number of packets sent by the port

Delta Port alive Duration (S) - The time port has been alive in seconds

Delta Packets Rx Dropped - Number of packets dropped by the receiver

Delta Packets Tx Dropped - Number of packets dropped by the sender

Delta Packets Rx Errors - Number of receive errors

Delta Packets Tx Errors - Number of transmit errors

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.

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

Checkis_valit - Indicates whether this load was built on valid values.

vpn_keyTable ID - Returns the Table ID values.

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.

Max Size - Returns the maximum size of this table.

Label - Label types for intrusions

Sample submission: You should submit a CSV file with a header row and the sample submission can be found below: