Network Security

Dataset KDD99 Network Intrusion Detection

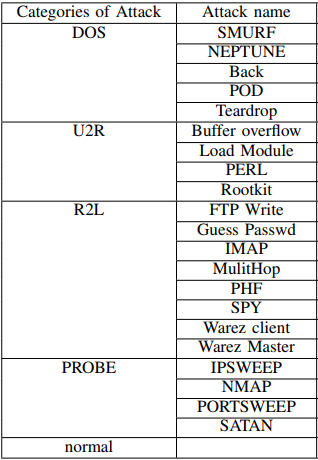


Figure Attack Categories

Figure Attack Categories

Dataset Description

1. Basic features of individual TCP connections.

|  |  |  |
| --- | --- | --- |
| *feature name* | *description* | *type* |
| duration | length (number of seconds) of the connection | continuous |
| protocol\_type | type of the protocol, e.g. tcp, udp, etc. | discrete |
| service | network service on the destination, e.g., http, telnet, etc. | discrete |
| src\_bytes | number of data bytes from source to destination | continuous |
| dst\_bytes | number of data bytes from destination to source | continuous |
| flag | normal or error status of the connection | discrete |
| land | 1 if connection is from/to the same host/port; 0 otherwise | discrete |
| wrong\_fragment | number of ``wrong'' fragments | continuous |
| urgent | number of urgent packets | continuous |

1. Content features within a connection suggested by domain

|  |  |  |
| --- | --- | --- |
| *feature name* | *description* | *type* |
| hot | number of ``hot'' indicators | continuous |
| num\_failed\_logins | number of failed login attempts | continuous |
| logged\_in | 1 if successfully logged in; 0 otherwise | discrete |
| num\_compromised | number of ``compromised'' conditions | continuous |
| root\_shell | 1 if root shell is obtained; 0 otherwise | discrete |
| su\_attempted | 1 if ``su root'' command attempted; 0 otherwise | discrete |
| num\_root | number of ``root'' accesses | continuous |
| num\_file\_creations | number of file creation operations | continuous |
| num\_shells | number of shell prompts | continuous |
| num\_access\_files | number of operations on access control files | continuous |
| num\_outbound\_cmds | number of outbound commands in an ftp session | continuous |
| is\_hot\_login | 1 if the login belongs to the ``hot'' list; 0 otherwise | discrete |
| is\_guest\_login | 1 if the login is a ``guest''login; 0 otherwise | discrete |

1. Traffic features computed using a two-second time window.

|  |  |  |
| --- | --- | --- |
| *feature name* | *description* | *type* |
| count | number of connections to the same host as the current connection in the past two seconds | continuous |
|  | *Note: The following  features refer to these same-host connections.* |  |
| serror\_rate | % of connections that have ``SYN'' errors | continuous |
| rerror\_rate | % of connections that have ``REJ'' errors | continuous |
| same\_srv\_rate | % of connections to the same service | continuous |
| diff\_srv\_rate | % of connections to different services | continuous |
| srv\_count | number of connections to the same service as the current connection in the past two seconds | continuous |
|  | *Note: The following features refer to these same-service connections.* |  |
| srv\_serror\_rate | % of connections that have ``SYN'' errors | continuous |
| srv\_rerror\_rate | % of connections that have ``REJ'' errors | continuous |
| srv\_diff\_host\_rate | % of connections to different hosts | continuous |

Steps:

1. Pre-processing data by removing redundant row values from the dataset.
2. Using Weka to determine accuracy by using correctly classified instances of each of the Machine-learning algorithm.
3. Coding the algorithm with the max precision and accuracy using R.
4. Further deploying that algorithm after parameter tuning for other dataset. (Testing)

Using Weka (Step 2)

**Naïve Bayes:**

Cross Validation (10 Folds)



Percentage Split: (70%)



**Decision Table:**

Cross Validation (10 Folds)



Percentage Split: (70%)



**KNN:**

Cross Validation (10 Folds)



Percentage Split: (70%)



**Random Forest:**

Cross Validation (10 Folds)



Percentage Split: (70%)



**AdaBoostM1**

Cross Validation (10 Folds)



Percentage Split: (70%)



**Attribute Selection:**