# KDD Cup 1999 Data Set

**Dataset files**

* <http://archive.ics.uci.edu/ml/machine-learning-databases/kddcup99-mld/kddcup.data_10_percent.gz>

**Dataset entry description**

Attribute 1 (duration) : Length (number of seconds) of the connection

Attribute 2 (protocol\_type) : type of the protocol, e.g. tcp, udp, etc.

Attribute 3 (service) : network service on the destination

Attribute 4 (flag) : normal or error state of the connection

Attribute 5 (src\_bytes) : number of data bytes from source to destination

Attribute 6 (dst\_bytes) : number of data bytes from destination to source

Attribute 7 (land) : 1 if connection is from/to the same host/port;

0 otherwise

Attribute 8 (wrong\_fragment) : number of “wrong” fragments

Attribute 9 (urgent) : number of urgent packets

Attribute 10 (hot) : number of “hot” indicators

Attribute 11 (num\_failed\_logins) : number of failed login attempts

Attribute 12 (logged\_in) : 1 if successfully logged in;

0 otherwise

Attribute 13 (num\_compromised) : number of “compromised” conditions

Attribute 14 (root\_shell) : 1 if root shell is obtained;

0 otherwise

Attribute 15 (su\_attempted) : 1 if ``su root'' command attempted;

0 otherwise

Attribute 16 (num\_root) : number of ``root'' accesses

Attribute 17 (num\_file\_creations) : number of file creation operations

Attribute 18 (num\_shells) : number of shell prompts

Attribute 19 (num\_access\_files) : number of operations on access control files

Attribute 20 (num\_outbound\_cmds) : number of outbound commands in an ftp session

Attribute 21 (is\_host\_login) : 1 if the login belongs to the “hot” list;

0 otherwise

Attribute 22 (is\_guest\_login) : 1 if the login is a “guest” login;

0 otherwise

Attribute 23 (count) : number of connections to the same host as the

current connection in the past two seconds

Attribute 24 (srv\_count) : number of connections to the same service as the

current connection in the past two seconds

Attribute 25 (serror\_rate) : % of connections that have “SYN” errors

Attribute 26 (srv\_serror\_rate) : % of connections that have “SYN” errors

Attribute 27 (rerror\_rate) : % of connections that have “REJ” errors

Attribute 28 (srv\_rerror\_rate) : % of connections that have “REJ” errors

Attribute 29 (same\_srv\_rate) : % of connections to the same service

Attribute 30 (diff\_srv\_rate) : % of connections to different services

Attribute 31 (srv\_diff\_host\_rate) : % of connections to different hosts

Attribute 32 (dst\_host\_count) : Destination host count

Attribute 33 (dst\_host\_srv\_count) : Destination host service count

Attribute 34 (dst\_host\_same\_srv\_rate) : Destination host same

Attribute 35 (dst\_host\_diff\_srv\_rate) : Destination host different service rate

Attribute 36 (dst\_host\_same\_src\_port\_rate) : Destination host same source port rate

Attribute 37 (dst\_host\_srv\_diff\_host\_rate) : Destination host service different host rate

Attribute 38 (dst\_host\_serror\_rate) : Destination host error rate

Attribute 39 (dst\_host\_srv\_serror\_rate) : Destination host service error rate

Attribute 40 (dst\_host\_rerror\_rate) : Destination host error rate

Attribute 41 (dst\_host\_srv\_rerror\_rate) : Destination service error rate

Attribute 42 (xAttack) : Attack type

**Labelling field**

Attribute 42 (xAttack)

Possible values (back dos, //

buffer\_overflow u2r, // writing data to a buffer until overruns the buffer's

boundary and overwrites adjacent memory locations

ftp\_write r2l, //

guess\_passwd r2l, // Password bruteforce attempt

imap r2l,

ipsweep probe,

land dos, // Flooding spoofed packet with the SYN flag

loadmodule u2r,

multihop r2l,

neptune dos, // SYN flood Denial of Service

nmap probe, // Nmap scan

perl u2r,

phf r2l, // phf cgi injection/code execution

pod dos, // ping of death

portsweep probe, // Port sweep attempt

rootkit u2r, // Unauthorized user access

satan probe, // Satan scan

smurf dos, // ICMP packet flooding using

spy r2l,

teardrop dos, // fragmented packets flooding

warezclient r2l,

warezmaster r2l)

**Example entry:**

61, tcp, telnet, SF, 294, 3929, 0, 0, 0, 0, 0, 1, 0, 1, 0, 4, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 255, 4, 0.02, 0.02, 0, 0, 0, 0.25, 0.73, 0.25, rootkit.

**Questions to answer using ML**

1. Rootkit is present on the network ? if yes then on which service rootkit is running.
2. Network is compromised by how many types of attacks?
3. Find the DOS duration on the network
4. Detected any nmap scan requests?
5. Which service is being used for DOS?

**Reference:**

* <http://archive.ics.uci.edu/ml/machine-learning-databases/kddcup99-mld/task.html>