Data Dictionary

KDD Data Set

The NSL-KDD data set with 42 attributes is used in this empirical study. This data set is an improvement over KDD'99 data set from which duplicate instances were removed to get rid of biased classification results. This data set has a number of versions available, out of which 20% of the training data is used which is identified as KDDTrain+_20Percent with a total number of 25192 instances. The test data set is identified by the name KDDTest+ and has a total of 22544 instances. Different configurations of this data set are available with variation in number of instances but the number of attributes in each case is 42. The attribute labeled 42 in the data set is the 'class' attribute which indicates whether a given instance is a normal connection instance or an attack. Table 1 gives the description of KDD data set attributes with class labels. Out of these 42 attributes, 41 attributes can be classified into four different classes as discussed below:

- Basic (B) Features are the attributes of individual TCP connections
- Content (C) features are the attributes within a connection suggested by the domain knowledge
- Traffic (T) features are the attributes computed using a two-second time window
- Host (H) features are the attributes designed to assess attacks which last for more than two seconds

Class wise detail of KDD data set attributes

S.No	Label	Attribute Name	S.No	Label	Attribute Name	S.No	Label	Attribute Name	S.No	Label	Attribute Name
1	B	duration	10	C	hot	23	T	count	32	H	dst_host_count
2	B	protocol_type	11	C	num_failed_logins	24	T	serror_rate	33	H	dst_host_srv_count
3	B	service	12	C	logged_in	25	T	rerror_rate	34	H	dst_host_same_srv_rate
4	B	src_bytes	13	C	num_compromised	26	T	same_srv_rate	35	H	dst_host_diff_srv_rate
5	B	dst_bytes	14	C	root_shell	27	T	diff_srv_rate	36	H	dst_host_same_src_port_rate
6	B	flag	15	C	su_attempted	28	T	srv_count	37	H	dst_host_srv_diff_host_rate
7	B	land	16	C	num_root	29	T	srv_serror_rate	38	H	dst_host_serror_rate
8	B	wrong_fragment	17	C	num_file_creations	30	T	srv_rerror_rate	39	H	dst_host_srv_serror_rate
9	B	urgent	18	C	num_shells	31	T	srv_diff_host_rate	40	H	dst_host_rerror_rate
			19	C	num_access_files				41	H	dst_host_srv_rerror_rate
			20	C	num_outbound_cmds				42	-	class
			21	C	is_hot_login						
			22	C	is_guest_login						