INTRUSION DETECTOR LEARNING

Software to detect network intrusions protects a computer network from unauthorized users, including perhaps insiders. The intrusion detector learning task is to build a predictive model (i.e. a classifier) capable of distinguishing between "bad" connections, called intrusions or attacks, and "good" normal connections.

A connection is a sequence of TCP packets starting and ending at some well-defined times, between which data flows to and from a source IP address to a target IP address under some well-defined protocol. Each connection is labeled as either normal, or as an attack.

FEATURES DESCRIPTION

feature name	*description*	*Type*
X	ID	discrete
duration	length (number of seconds) of the connection	continuous
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

Table 1: Basic features of individual TCP connections.

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	continuous

	access control files	
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

Table 2: Content features within a connection suggested by domain knowledge.

feature name	*description*	*type*
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
diff_srv_rate	% of connections to different services	continuous
Note: The following features refer to these same-service connections.		
srv_rerror_rate	% of connections that have “REJ” errors	continuous
srv_diff_host_rate	% of connections to different hosts	continuous
dst_host_count	-	continuous
dst_host_srv_diff_host_rate	-	continuous
dst_host_rerror_rate	-	continuous
dst_host_srv_rerror_rate	-	continuous
Target	0 – Normal; 1 - Intrusion	discrete

Table 3: Traffic features computed using a two-second time window.

You will be evaluated on the following parameters:

□ Data Pre-processing.

□ Visualizations & data exploration for meaningful insights.

□ Coding standards. This includes how concise & well commented the code is.

□ Performance of the final algorithm.