Case Study on Network Intrusion Detection

CASE STUDY

Background

The dataset to be audited was provided which consists of a wide variety of intrusions simulated in a military network environment. It created an environment to acquire raw TCP/IP dump data for a network by simulating a typical US Air Force LAN. The LAN was focused like a real environment and blasted with multiple attacks. A connection is a sequence of TCP packets starting and ending at some time duration between which data flows to and from a source IP address to a target IP address under some well-defined protocol. Also, each connection is labeled as either normal or as an attack with exactly one specific attack type. Each connection record consists of about 100 bytes.

For each TCP/IP connection,41 quantitative and qualitative features are obtained from normal and attack data (3 qualitative and 38 quantitative features) .The class variable has two categories:

- Normal
- Anomalous

Objective:

1. Kindly Work on Exploratory Data Analysis
2. Build multiple models to classify each connection as either normal or anomalous. Use the Train data set to train the model and use that model to predict the response variable value in the Test data set.

Data basically represents the packet data for a time duration of 2 seconds.

1-9 Columns: basic features of packet (type 1)

10-22 columns: employ the content features (type 2)

23-31 columns: employ the traffic features with 2 seconds of time window (type 4)

32-41 columns: employ the host based features

C: Continuous data

D: Discrete data

Feature name	Variable type	Type	Description
Duration	C	1	No. of seconds of the connection
Protocol_type	D	1	Type of protocol E.g.: TCP,UDP ,ICMP

Service	D	1	Network service on the destination E.g.:http,telnet
Flag	D	1	Normal or error status of the connection
src_bytes	C	1	Number of data bytes from source to destination
dst_bytes	C	1	Number of data bytes from destination to source
Land	D	1	1-connection is from the same host/port: 0-otherwise
Wrong_fragment	C	1	No. of 'wrong' fragments
Urgent	C	1	No of urgent fragments
Hot	C	2	The count of access to system directories, creation and execution of programs
Num_failed_logins	C	2	No. of failed login attempts
Logged_in	D	2	1-successfully logged in 0-otherwise
num_compromised	C	2	No. of compromised conditions
Root_shell	C	2	1-root shell is obtained;0 otherwise
Su_attempted	C	2	1-'su root' command attempted;0 otherwise
Num_root	C	2	No .of root accesses
num_file_creations	C	2	Number of file creation operations
Num_shells	C	2	No of shell prompts
Num_access_files	C	2	No. of write ,delete and create operations on access control files
Num_outbound_cmds	C	2	No. of outbound commands in an ftp session
Is_hot_login	D	2	1-the login belongs to the 'hot' list 0: otherwise
Count	C	3	No. of connections to the same host as the current connection in the past seconds
Srv_count	C	3	No of connections to the same host as the current connection in the past 2 seconds
serror_rate	C	3	% of connections that have 'SYN' errors to the same host
Srv_serror_rate	C	3	% of connections that have 'SYN' errors to the same service

Rerror_rate	C	3	% of connections that have 'REJ' errors to the same host
Srv_diff_host_rate	C	3	% of connections to different services and to the same host
Dst_host_count	C	3	No of connections to the same host to the destination host as the current connection in the past 2 seconds
Dst_host_srv_count	C	3	No of connections from the same service to the destination host as the current connection in the past 2 seconds
dst_host_srv_count	C	3	No. of connections from the same service to the destination host as the current connection in the past 2 seconds
Dst_host_srv_count	C	3	No. of connections from the same service to the destination host as the current connection in the past 2 seconds
Dst_host_same_srv_rate	C	3	% of connections from the same service to the destination host
Dst_host_diff_srv_rate	C	3	% of connections from the different services to the destination host
Dst_host_same_src_port_rate	C	3	% of connections from the port services to the destination host
Dst_host_srv_diff_host_rate	C	3	% of connections from the different hosts from the same service to destination host
Dst_host_serror_rate	C	3	% of connections that have 'SYN" errors to same host to the destination host
dst_host_srv_serror_rate	C	3	% of connections that have 'SYN' errors from the same service to the destination host
Dst_host_rerror_rate	C	3	% of connections that have 'REJ' errors from the same host to destination host
Dst_host_srv_rerror_rate	C	3	% of connections that have 'REJ' errors from the same service to the destination host