Data Analytics in Cyber Security
CT115-3-M (Version E)

Assignment Briefing

Intrusion Lifecycle

Phase	Technique	Description
1	**Reconnaissance**	- Gather as much info about targets as possible. - Required to craft an attack.
2	**Initial exploitation**	- Gain access to network or hosts, obtain credentials, etc.
3	**Privilege escalation**	- Gain greater control over systems. - Can do more damage with higher privileges.
4	**Pivoting**	- Compromise a central host. - Spread to other hosts and network segments.
5	**Persistence**	- Maintaining access is an important goal. - Avoiding discovery, erasing traces of activity

Intrusion Detection Datasets

- 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.
- Intrusion Detection Datasets need to **tag patterns of activity**, **not individual instances** (like malware /spam/phishing/uploads).
- Intrusion Detection Systems are designed to monitor these patterns **using the limited information available in the network packet.**

KDD99 Dataset

- 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 record consists of about 100 bytes.
- Each connection is labeled as either **normal** or a specific **attack type**. The datasets contain a total of 22 training attack types, with an additional 17 types in the test data only.
- Attack types (exploits) fall into four **categories**:
  - **DOS**: denial-of-service, e.g., syn flood;
  - **Probe**: surveillance and other probing, e.g., port scanning;
  - **R2L**: remote-to-local, unauthorized access from a remote machine, e.g., guessing password;
  - **U2R**: user-to-root, unauthorized access to local superuser (root) privileges, e.g., various “buffer overflow" attacks.

KDD99 Dataset

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

Basic features of individual TCP connections.

KDD99 Dataset

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
Traffic features computed using a two-second time window.		

KDD99 Dataset

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
Content features within a connection suggested by domain knowledge.		