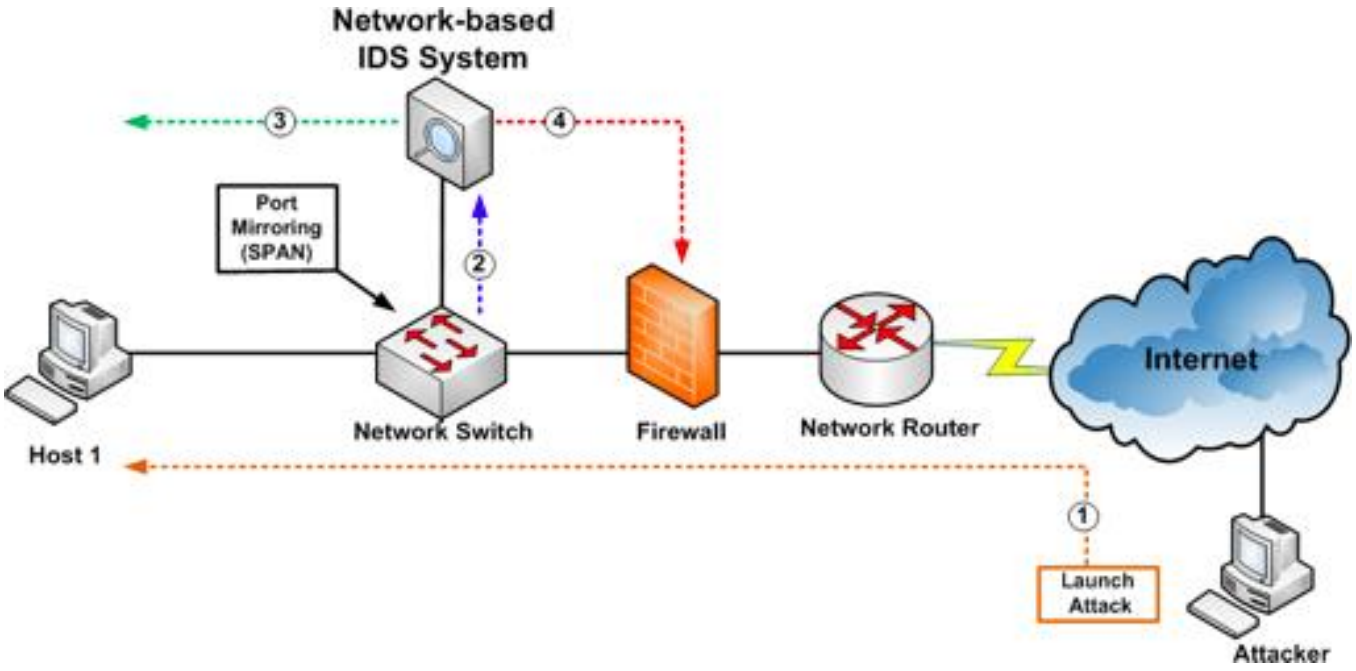
Cyber Security Attack Defender

Kaggle link

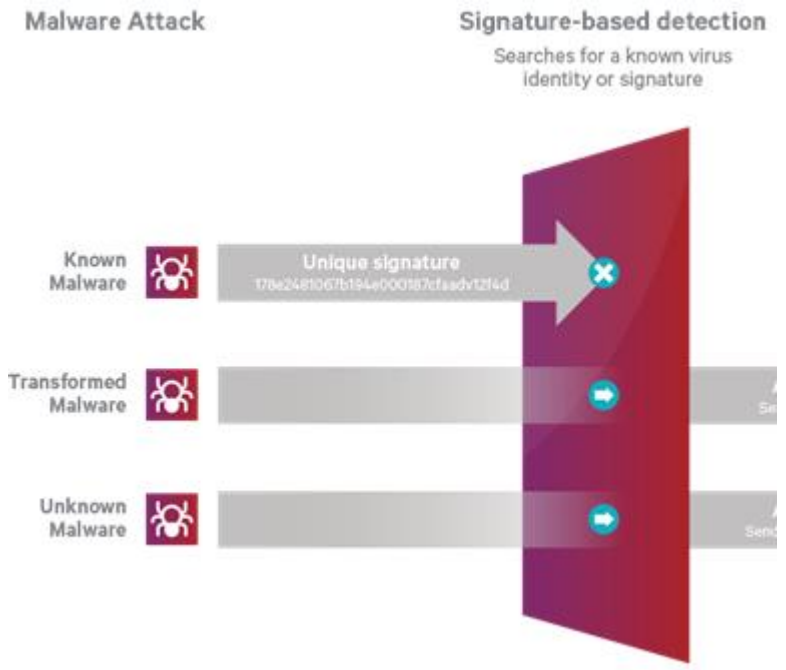
Introduction

Cyber Security Attack Defender



Traditional ways in Firewall

- Signature-based
- What if we use information of connection?



Dataset

- Based on DARPA'98 IDS evaluation program
- About 5 million connection records
- Number of columns: 42
  - 41 for connection information

Task

- Classify normal / bad connections
- Four types of attack
  - DOS: denial-of-service
  - R2L: unauthorized access from a remote machine
  - U2R: unauthorized access to local superuser (root) privileges
  - probing: surveillance and other probing

- Your Job is to predict the status of a connection :
  - Normal
  - What type of attack?

Columns Description

- Basic features
  - 9 columns
- Content features
  - 13 columns based on the action of the connections (ex. Login, sudo……)
- Traffic features (2-seconds window)
  - 19 columns
  - Same host：same destination as the current connection
  - Same service：same service as the current connection

Feature	Description	Type
1. duration	Duration of the connection.	Cont.
2. protocol type	Connection protocol (e.g. tcp, udp)	Disc.
3. service	Destination service (e.g. telnet, ftp)	Disc.
4. flag	Status flag of the connection	Disc.
5. source bytes	Bytes sent from source to destination	Cont.
6. destination bytes	Bytes sent from destination to source	Cont.
7. land	1 if connection is from/to the same host/port; 0 otherwise	Disc.
8. wrong fragment	number of wrong fragments	Cont.
9. urgent	number of urgent packets	Cont.
10. hot	number of "hot" indicators	Cont.
11. failed logins	number of failed logins	Cont.
12. logged in	1 if successfully logged in; 0 otherwise	Disc.
13. # compromised	number of "compromised" conditions	Cont.
14. root shell	1 if root shell is obtained; 0 otherwise	Cont.
15. su attempted	1 if "su root" command attempted; 0 otherwise	Cont.
16. # root	number of "root" accesses	Cont.
17. # file creations	number of file creation operations	Cont.
18. # shells	number of shell prompts	Cont.
19. # access files	number of operations on access control files	Cont.
20. # outbound cmds	number of outbound commands in an ftp session	Cont.
21. is hot login	1 if the login belongs to the "hot" list; 0 otherwise	Disc.
22. is guest login	1 if the login is a "guest" login; 0 otherwise	Disc.
23. Count	number of connections to the same host as the current connection in the past two seconds	Cont.
24. srv count	number of connections to the same service as the current connection in the past two seconds	Cont.
25. serror rate	% of connections that have "SYN" errors	Cont.
26. srv serror rate	% of connections that have "SYN" errors	Cont.
27. rerror rate	% of connections that have "REJ" errors	Cont.
28. srv rerror rate	% of connections that have "REJ" errors	Cont.
29. same srv rate	% of connections to the same service	Cont.
30. diff srv rate	% of connections to different services	Cont.
31. srv diff host rate	% of connections to different hosts	Cont.
32. dst host count	count of connections having the same destination host	Cont.
33. dst host srv count	count of connections having the same destination host and using the same service	Cont.
34. dst host same srv rate	% of connections having the same destination host and using the same service	Cont.
35. dst host diff srv rate	% of different services on the current host	Cont.
36. dst host same src port rate	% of connections to the current host having the same src port	Cont.
37. dst host srv diff host rate	% of connections to the same service coming from different hosts	Cont.
38. dst host serror rate	% of connections to the current host that have an S0 error	Cont.
39. dst host srv serror rate	% of connections to the current host and specified service that have an S0 error	Cont.
40. dst host rerror rate	% of connections to the current host that have an RST error	Cont.
41. dst host srv rerror rate	% of connections to the current host and specified service that have an RST error	Cont.

Attack types

- Four types of attacks included manifold methods of malware
- See training_attack_types.txt for details

Submission

- Please label the connection as follows:
  - 0 -> normal
  - 1 -> dos
  - 2 -> u2r
  - 3 -> r2l
  - 4 -> probe

For your reference…….

- Search for DARPA1998
- A survey of DM and ML methods for Cyber Security Intrusion Detection