Dataset Context

The dataset 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  
• Anomaly**

Columns Information

**Protocol:**

1. TCP:
2. UDCP:
3. ICMP:

**Service:**

1. Auth:
2. Exec:
3. Domain:
4. Courier:
5. KShell:
6. KLogin:
7. Http:

**Flag:**

Table

Description automatically generated

1. **Duration**: Time duration of the connection (quant.)
2. **Protocol type**: Protocol used in connection (cat., 3 categories)
3. **Service**: Destination network service used (cat., 70 categories)
4. **Flag**: status of the connection (e.g. **REJ** = connection rejected) (cat., 11 categories)
5. **Src bytes**: number of data bytes transferred from source to destination (quant.)
6. **Dst bytes**: number of data bytes transferred from destination to source (quant.)
7. **Land**: indicator whether port number and IP address of source and destination are equal, if yes = 1, otherwise 0 (binary)
8. **Wrong fragment**: number of wrong fragments in connection (quant.)
9. **Urgent**: number of urgent packets (quant.)
10. **Hot**: number of ”hot” indicators in the content such as: entering a system directory, creating programs and executing programs (quant.)
11. Num failed logins: number of failed login attempts (quant.)
12. logged in: 1 if successfully logged in, 0 otherwise (binary)
13. num compromised: number of ”compromised” conditions (quant.)
14. root shell: 1 if root shell is obtained, 0 otherwise (binary)
15. su attempted: 1 if ”su root” command attempted or used, 0 otherwise (quant., data set contains value 2)
16. num root: number of operations performed as a root or root accesses (quant.)
17. num file creations: number of file creation operations (quant.)
18. num shells: number of shell prompts (quant.) 51 3. Data NSL-KDD’99
19. num access files: number of operations on access control files (quant.)
20. num outbound cmds: number of outbound commands in an ftp session (quant.)
21. is host login: 1 if the login is from root or admin, 0 otherwise (binary)
22. is guest login: 1 if the login is from guest, 0 otherwise (binary)
23. count: number of connections to the same destination host as the current connection in the past 2 seconds (quant.)
24. srv count: number of connections to the same service (port number) as the current connection in the past 2 seconds (quant.)
25. serror rate: % of connections that have activated s0, s1, s2 or s3 flag (4) among connections aggregated in count (quant.) 52 3. Data NSL-KDD’99
26. srv serror rate: % of connections that have activated s0, s1, s2 or s3 flag (4) among connections aggregated in srv count (quant.)
27. rerror rate: % of connections that have activated REJ flag (4) among connections aggregated in count (quant.)
28. srv rerror rate: % of connections that have activated REJ flag (4) among connections aggregated in srv count (quant.)
29. same srv rate: % of connections to the same service among those aggregated in count (quant.)
30. diff srv rate: % of connections to the different service among those aggregated in count (quant.)
31. srv diff host rate: % of connections that were to different destination machines among the connections aggregated in srv count (quant.)
32. dst host count: count of the connections having the same destination IP address (quant.)
33. dst host srv count: count of connections having the same port number (quant.)
34. dst host same srv rate: % of connections that were to different services, among those in dst host count (quant.)
35. dst host diff srv rate: % of connections that were to different services, among those in dst host count (quant.)
36. dst host same src port rate: % of connections that were to the same source port, among those in dst host srv count (quant.)
37. dst host srv diff host rate: % of connections that were to different destination machines, among those in dst host srv count (quant.)
38. dst host serror rate: % of connections that have activated the s0, s1, s2 or s3 flag (4), among those in dst host count (quant.)
39. dst host srv serror rate: % of connections that have activated the s0, s1, s2 or s3 flag (4), among those in dst host srv count (quant.)
40. dst host rerror rate: % of connections that have activated the REJ flag (4), among those in dst host count (quant.)
41. dst host srv rerror rate: % of connections that have activated the REJ flag (4), among those in dst host srv count (quant.)
42. **class**: If the connection is Normal, or Anormaly.