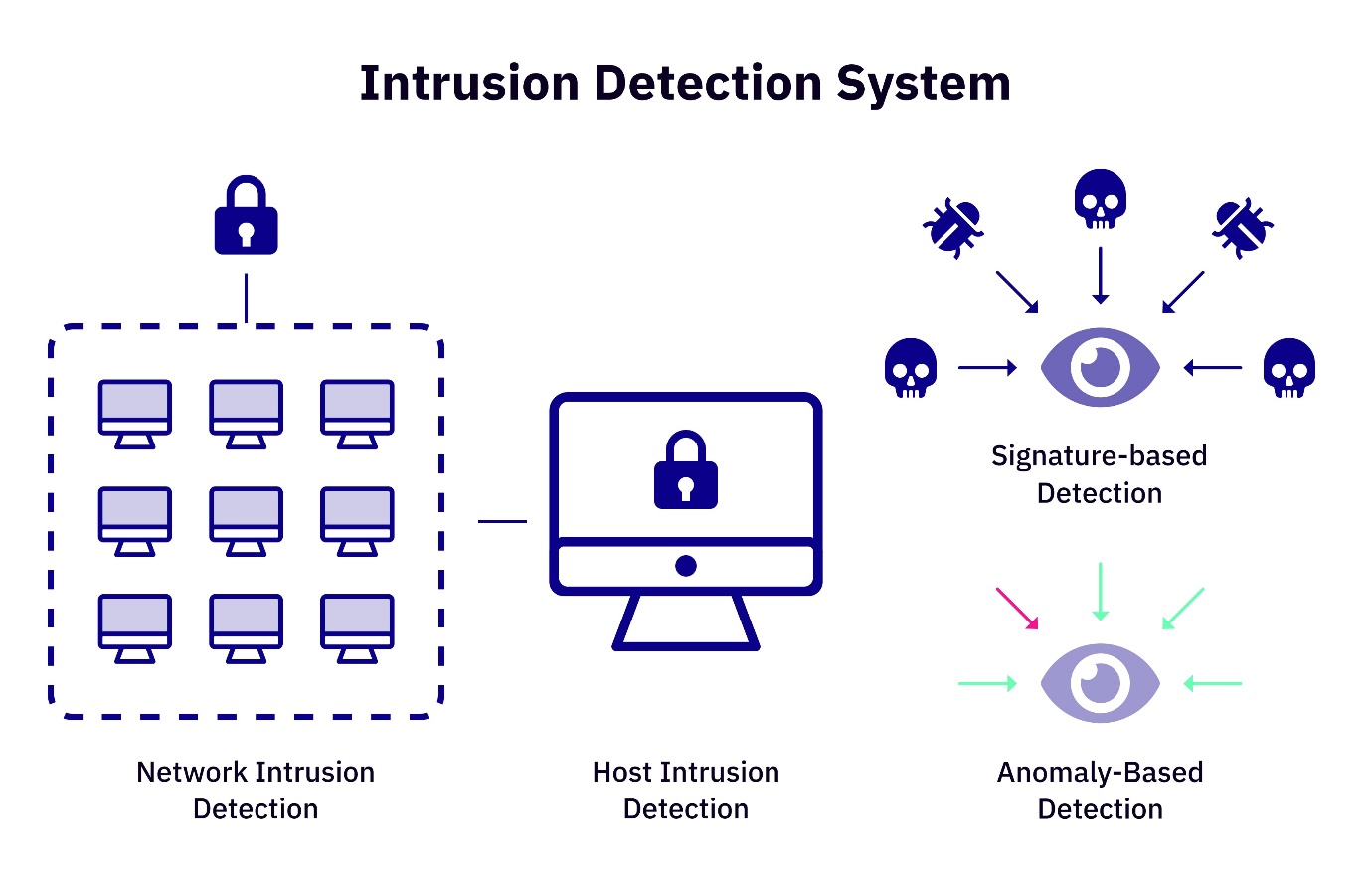
**INTRUSION DETECTION SYSTEM USING ML**



**Software Used:**

* **Jupyter Notebook**
* **Data-set source** [**https://www.unb.ca/cic/datasets/ids-2017.html**](https://www.unb.ca/cic/datasets/ids-2017.html)

We are making a ML Intrusion detection system, using ML algorithms and concepts, we try to make this project extend in other way by making a web application or any Network monitoring host further.

**Introduction:**

An Intrusion Detection System (IDS) is a network security technology originally built for detecting vulnerability exploits against a target application or computer. The IDS is also a **listen-only** device. The IDS monitors traffic and reports results to an administrator. It cannot automatically take action to prevent a detected exploit from taking over the system.

In this ML project report, we are focusing of network attributes and their characteristics:

* **Duration**: The length of time for which the connection was active (in seconds). Unusually long or short connections, which might be indicative of suspicious activity.
* **Type of Protocol**: The network protocol type used in the connection. Protocols used in this dataset are **TCP, ICMP, UDP.**
* **Service**: The network service on the destination host. There are many services present in the dataset.

**Example**: http, pop\_2, pop\_3, login, shell, exec, ftp, smtp etc.

* **flag**: Flags associated with the connection. This shows the connection standards like ACK, SYN-ACK, FIN, REG, SF etc.

Flags are analyzed to detect certain types of attacks, such as port scanning or SYN flooding.

**Here are details of few flags found in this ML data-set are:**

**REJ**: "REJ" indicates that a connection request was rejected, often due to a firewall rule or other security measures. The connection was explicitly denied.

**SF**: "SF" stands for "SYN/ACK - FIN." It indicates a normal, established connection where a TCP handshake was successfully completed.

**RSTO**: "RSTO" signifies that the connection was reset by the originator. This might happen when a connection attempt is made to a non-listening port.

**S0**: "S0" denotes that a connection attempt was made but no response was received. It's often associated with a closed or filtered port.

* **Source** **bytes**: The number of data bytes sent from the source to the destination.
* **Destination** **bytes**: The number of data bytes sent from the destination to the source.

data exfiltration or large data transfers that might be indicative of an attack.

* **Land**: The "land" attribute indicates whether the source and destination IP addresses and ports are the same. This can be used to detect IP spoofing or certain types of attacks that involve misleading source and destination information.
* **Wrong fragments**: The number of wrong fragments in a connection attempt.
* **Urgent**: The number of urgent packets in the connection.
* **hot**: The number of "hot" indicators in the content of the data.
* **Failed** **logins**: The number of failed login attempts.
* **Logged in:** Indicates whether the user is logged in (1 if yes, 0 if no).

"Compromised conditions" typically refers to situations where a system, network, or resource has been successfully breached or compromised by an unauthorized user or an attacker.

* **Compromised conditions**: The number of compromised conditions.
* **Root shell:** Indicates whether a root shell is obtained (1 if yes, 0 if no).
* **Attempted** “**SU**”: Indicates whether "su" root command attempted (1 if yes, 0 if no).

"su" stands for "**Superuser**" or "**Switch** **User**." It is a command used to gain elevated privileges, allowing you to run commands as the root user or another user with superuser privileges.

There are many attributes related to shells, commands, files in machines either Linux, Windows or other.

* **num\_root**: The number of "root" accesses.
* **num\_file\_creations**: The number of file creations.
* **num\_shells**: The number of shell prompts.
* **num\_access\_files**: The number of access control files.
* **num\_outbound\_cmds**: The number of outbound commands.
* **is\_host\_login**: Indicates whether the login belongs to the "host\_login" class (1 if yes, 0 if no).
* **is\_guest\_login**: Indicates whether the login is a "guest" login (1 if yes, 0 if no).
* **count**: The number of connections to the same host as the current connection.

which can be used to detect scanning or reconnaissance activities.

* **srv\_count**: The number of connections to the same service as the current connection.

There are other attributes related to rates and here are the attributes:

* **serror\_rate**: The percentage of connections that have "SYN" errors.
* **srv\_serror\_rate**: The percentage of connections to the same service that have "SYN" errors.
* **rerror\_rate:** The percentage of connections that have "REJ" errors.
* **srv\_rerror\_rate**: The percentage of connections to the same service that have "REJ" errors.
* **same\_srv\_rate**: The percentage of connections to the same service.
* **diff\_srv\_rate**: The percentage of connections to different services.
* **srv\_diff\_host\_rate**: The percentage of connections to different hosts among the same services.
* **dst\_host\_count**: The number of connections to the same destination host.
* **dst\_host\_srv\_count**: The number of connections to the same service on the destination host.
* **dst\_host\_same\_srv\_rate**: The percentage of connections to the same service on the destination host.
* **dst\_host\_diff\_srv\_rate**: The percentage of connections to different services on the destination host.
* **dst\_host\_same\_src\_port\_rate**: The percentage of connections from the same source port.
* **dst\_host\_srv\_diff\_host\_rate**: The percentage of connections to the same service but to different hosts on the destination host.
* **dst\_host\_serror\_rate**: The percentage of connections to the same destination host that have "SYN" errors.
* **dst\_host\_srv\_serror\_rate**: The percentage of connections to the same service on the destination host that have "SYN" errors.
* **dst\_host\_rerror\_rate**: The percentage of connections to the same destination host that have "REJ" errors.
* **dst\_host\_srv\_rerror\_rate**: The percentage of connections to the same service on the destination host that have "REJ" errors.

These attributes provide information about the error rates in connection attempts and can help identify connections with a high error rate, which may indicate a suspicious activity.

**References:**

<https://www.paloaltonetworks.com/cyberpedia/what-is-an-intrusion-detection-system-ids#:~:text=An%20Intrusion%20Detection%20System%20(IDS,reports%20results%20to%20an%20administrator>.