1. Project Goals Summary

The **SentinelNet project** aims to build an intelligent system that can detect, classify, and respond to cybersecurity threats in real time. Its primary goal is to leverage **AI and machine learning** to analyze network traffic, identify malicious patterns, and differentiate between normal and abnormal activities. Unlike traditional security systems that rely on fixed rules, SentinelNet focuses on **adaptive defense mechanisms**, meaning it can learn from new attack patterns and continuously improve detection. In simple terms, SentinelNet is about creating a **smart cyber guardian** that safeguards networks against both known and emerging threats.

---

2. Dataset Exploration

To build and evaluate SentinelNet, we rely on standard benchmark datasets that capture both benign traffic (normal user activity) and malicious traffic (cyberattacks). Two widely used datasets are NSL-KDD and CICIDS2017.

---

1. NSL-KDD Dataset

The **NSL-KDD** dataset is an improved version of the earlier **KDD Cup 1999 dataset**, designed for network intrusion detection research.

- **Number of Features:**
  - 41 input features (network traffic characteristics).
  - 1 label feature (attack type/normal).
- **Feature Types:**
  - **Basic Features:** Protocol type, duration, service, flag, etc.
  - **Content Features:** Failed login attempts, shell access, number of file creations, etc.
  - **Traffic Features:** Count of connections to the same host, same service, etc.
- **Attack Categories:**
  - **Denial of Service (DoS):** Attackers flood resources (e.g., Smurf, Neptune).
  - **Probe:** Scanning to gather information (e.g., Nmap, Satan).
  - **User-to-Root (U2R):** Unauthorized root access (e.g., buffer overflow).
  - **Remote-to-Local (R2L):** Unauthorized local access via remote system (e.g., guess-password).
- **Strengths:** Cleaner than KDD'99 (duplicate and redundant records removed).
- **Weaknesses:** Outdated; doesn't fully represent modern attacks.

---

2. CICIDS2017 Dataset

The **CICIDS2017** dataset, created by the Canadian Institute for Cybersecurity, is one of the most realistic modern intrusion detection datasets. It simulates actual enterprise network traffic over five days, including both **normal behavior** and **contemporary attacks**.

- **Number of Features:**
    - More than 80 statistical features extracted from packet flows.
    - Examples: Flow duration, packet length, inter-arrival time, forward/backward packet counts, header flags.
- **Attack Types Covered:**
    - **Denial of Service (DoS) & Distributed DoS (DDoS):** Flooding resources with traffic.
    - **Brute Force Attacks:** Password guessing on SSH and FTP services.
    - **Web-based Attacks:** SQL injection, XSS (Cross-site scripting).
    - **Infiltration:** Malware inserted into a network from inside.
    - **Botnet:** Compromised systems controlled remotely.
    - **Heartbleed:** Exploiting OpenSSL vulnerability.
- **Strengths:** Modern, diverse, realistic traffic with labeled attacks.
- **Weaknesses:** Larger and more complex; requires high computational resources for analysis.

---

3. Key Differences Between NSL-KDD and CICIDS2017

Aspect	NSL-KDD	CICIDS2017
Year	2009 (based on KDD'99 - 1999)	2017
Features	41	80+
Attack Categories	4 broad categories (DoS, Probe, U2R, R2L)	Multiple modern categories (DoS/DDoS, Brute Force, Web attacks, Botnet, Infiltration, Heartbleed)
Traffic Nature	Synthetic, simulated	Realistic enterprise traffic
Size	Smaller, easier to handle	Very large, computationally demanding
Usefulness	Good for learning basics, benchmarking	Best for modern, real-world IDS research