# Numerical Discussion

## Edge IIOT Dataset

The **Edge-IIoTset dataset** is built upon a sophisticated **seven-layer testbed architecture**, integrating diverse technologies such as ThingsBoard, OPNFV, Hyperledger Sawtooth, Digital Twin, ONOS SDN controller, and Mosquitto MQTT brokers. This comprehensive design facilitates the generation of IoT data from **more than 10 types of IoT devices**, including various sensors for temperature, humidity, water level, pH, soil moisture, heart rate, and flame detection.

The dataset meticulously identifies and analyzes **14 specific attack types**, which are categorized into **five major cybersecurity threats**: DoS/DDoS, information gathering, Man-in-the-Middle, injection, and malware attacks. It is a large-scale collection, featuring **over 20 million labeled samples**. From an initial pool of **1176 features**, the dataset was refined to include **61 highly correlated features**, extracted from diverse sources such as alerts, system resources, logs, and network traffic. This rich, multi-dimensional data supports the development and evaluation of machine learning-based intrusion detection systems in both centralized and federated learning paradigms, making it a robust resource for advanced cybersecurity research.

## Ton_Iot Dataset

The **TON_IoT dataset** is a comprehensive, cross-domain collection of telemetry, operating system, and network-level data, specifically designed for AI-driven cybersecurity in IoT and Industrial IoT (IIoT) systems. Its network stream subset alone comprises **over 21 million samples**. A key subset used for IoT intrusion detection contains **461,043 IoT traffic entries**, which are distinctly divided into **300,000 normal** and **161,043 attack records**.

This rich dataset features **44 distinct features** relevant to network and system behavior and encompasses **nine types of anomalies**, including XSS, DDoS, DoS, scanning, injection, ransomware, backdoor, password, and Man-in-the-Middle (MITM) attacks. It is commonly partitioned with an **80:20 ratio for training and testing**. Performance evaluations on this dataset demonstrate its robust utility: a GAN-Transformer framework achieved average precision, recall, and F1 scores of **97.62%, 97.66%, and 97.64%** respectively. Furthermore, an optimized hybrid deep learning model, Transformer–GAN–AE, showcased impressive results with **98.92% accuracy, 99.52% recall, and a 99.87% Area Under the Curve (AUC)**, validating its capability to benchmark advanced intrusion detection systems effectively.