Generating benign network datasets from IoT devices for the training of AI-driven intrusion detection systems

In a recent study, we investigated the potential of Large Language Models (LLMs) to augment IoT Intrusion Detection System (IDS) datasets, achieving promising results. A key prerequisite for training such AI systems is collecting authentic, labelled network data from IoT devices.

The task for this topic is to build an IoT test bed comprising five or more IoT devices and to capture their network traffic behaviour at a central observation point. The aim is to capture high-quality packet-level data from real devices in use and extract relevant features in a human-readable taxonomy. This could then be used to train and evaluate an LLM-driven synthetic traffic generator and an intrusion detection framework.

To work successfully on this topic, you should have a basic understanding of IoT devices and be familiar with computer network basics such as the components of the ISO-OSI (TCP/IP) network reference model, common protocols and topologies such as Ethernet and Wi-Fi, and network packets and flows. You should also be willing to learn about smart home devices and their functionality quickly.

Note: This represents only a preliminary high-level description of the topic. Details and actual scoping of the thesis will be agreed on with separately with the supervisor.

Thesis supervisor:
Prof. Dr.-Ing. Markus Miettinen
markus.miettinen@fra-uas.de
+49 69 1533 3969

Secondary supervisor: Maurizio Petrozziello
petrozziello@fra-uas.de
+49 69 1533 3673

Thesis language: English or German

Topic updated: June 11, 2025