

Platform Variability in Edge-Cloud Vision Systems

Ali J. Ben Ali (alijmabe@buffalo.edu), Sofiya Semenova (sofiyase@buffalo.edu), Karthik Dantu (kdantu@buffalo.edu)

Distributed Robotics and Networked Embedded Systems (DRONES) Lab

Department of Computer Science and Engineering | University at Buffalo

MODERN EDGE DEVICES

- Good computing (CPU, GPU, custom processing)
- Multiple interaction modalities (touch, voice, gesture)
- Various connectivity modalities (Wi-Fi, cellular, Bluetooth, NFC)
- Lots of sensing including cameras, depth sensors, inertial sensors and others
- Mobility and AR/MR/VR allow for richer set of applications to be deployed



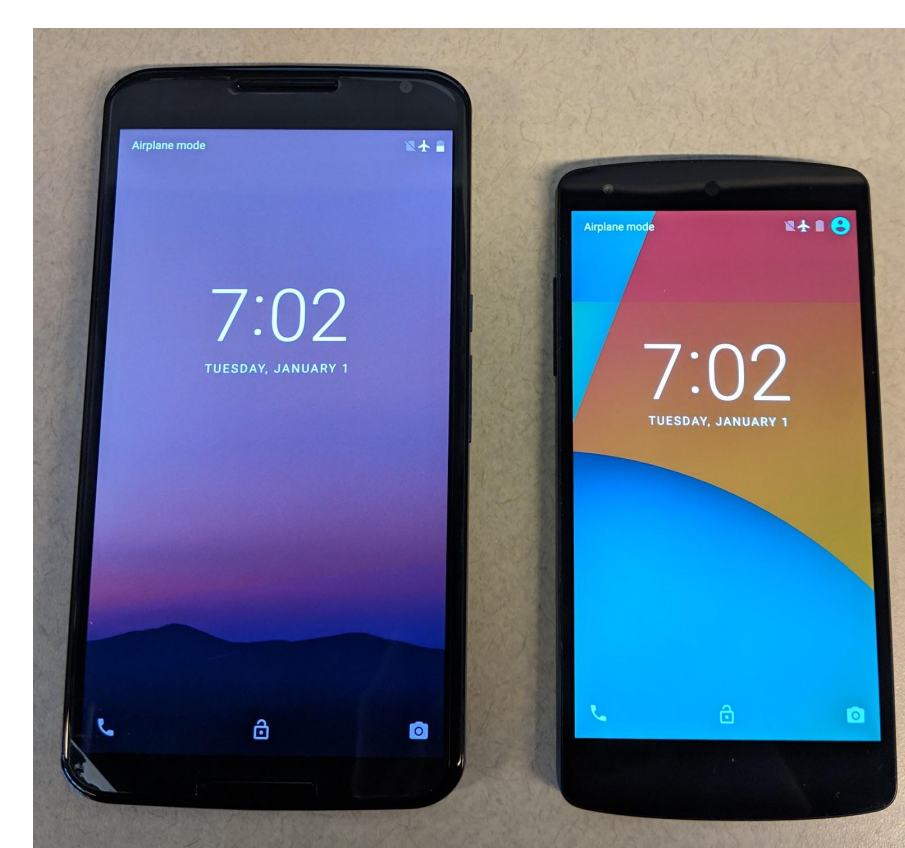
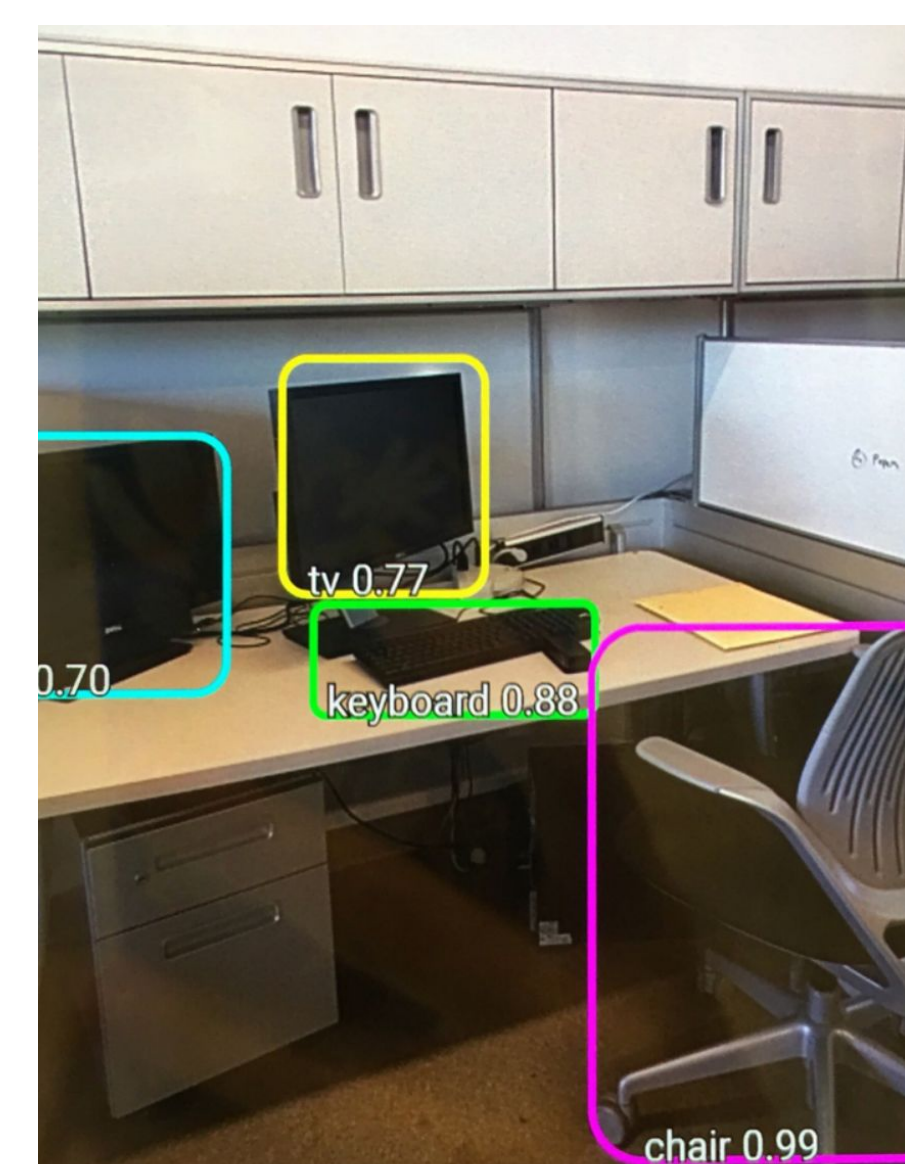
APPLICATIONS

- **Immersive Gaming:** combines realistic scenarios with virtual elements while gaming such as Pokemon Go
- **Driving Navigation Dashboard:** allows seamless visualization of details of interest reducing driver distraction
- **Digital Manufacturing:** integrates digital design, manufacturing, and maintenance to simplify product lifecycle
- **Precision Agriculture:** smart sensing used to adapt watering, fertilizer, and pesticide

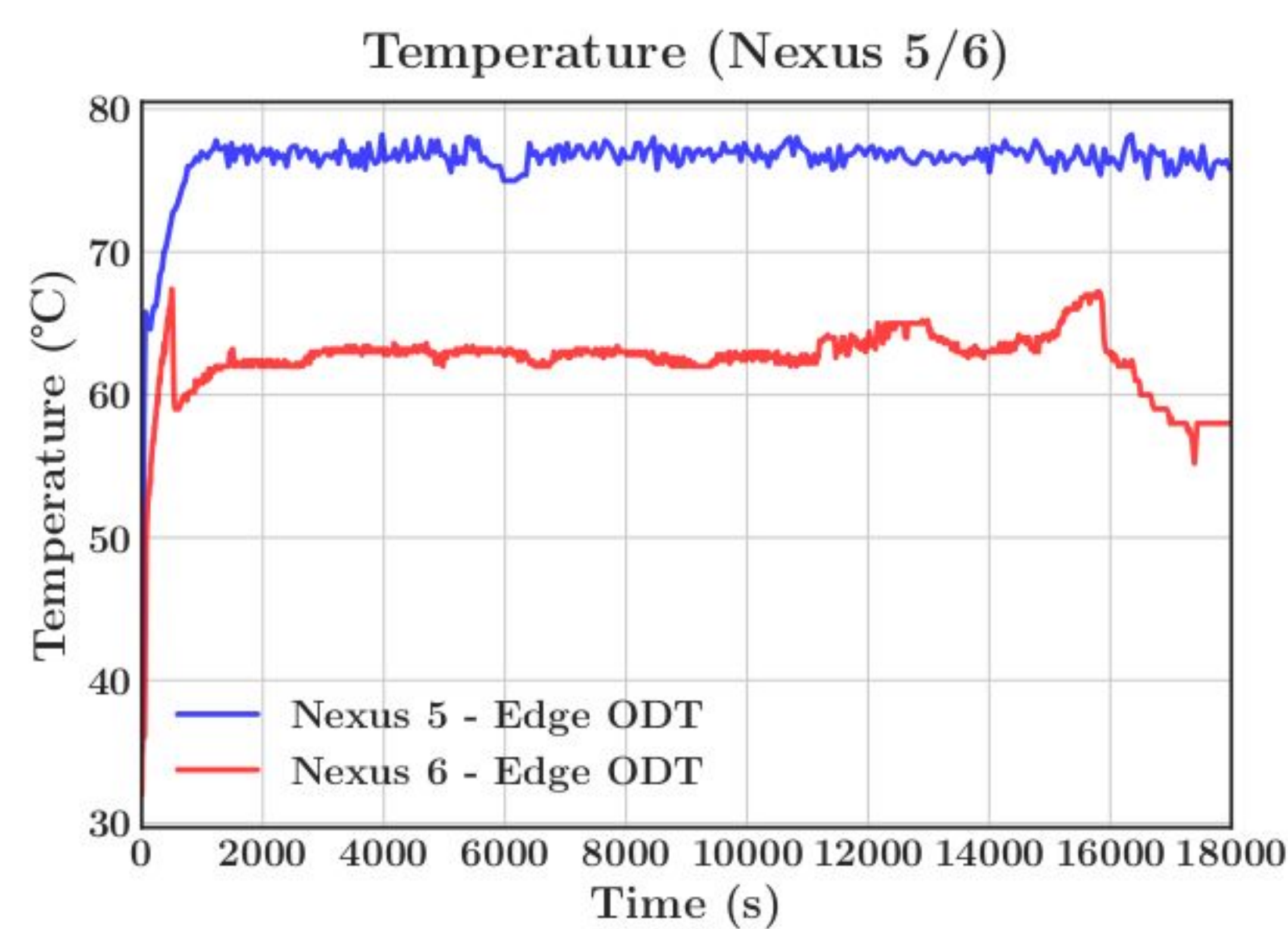


STUDY SETUP

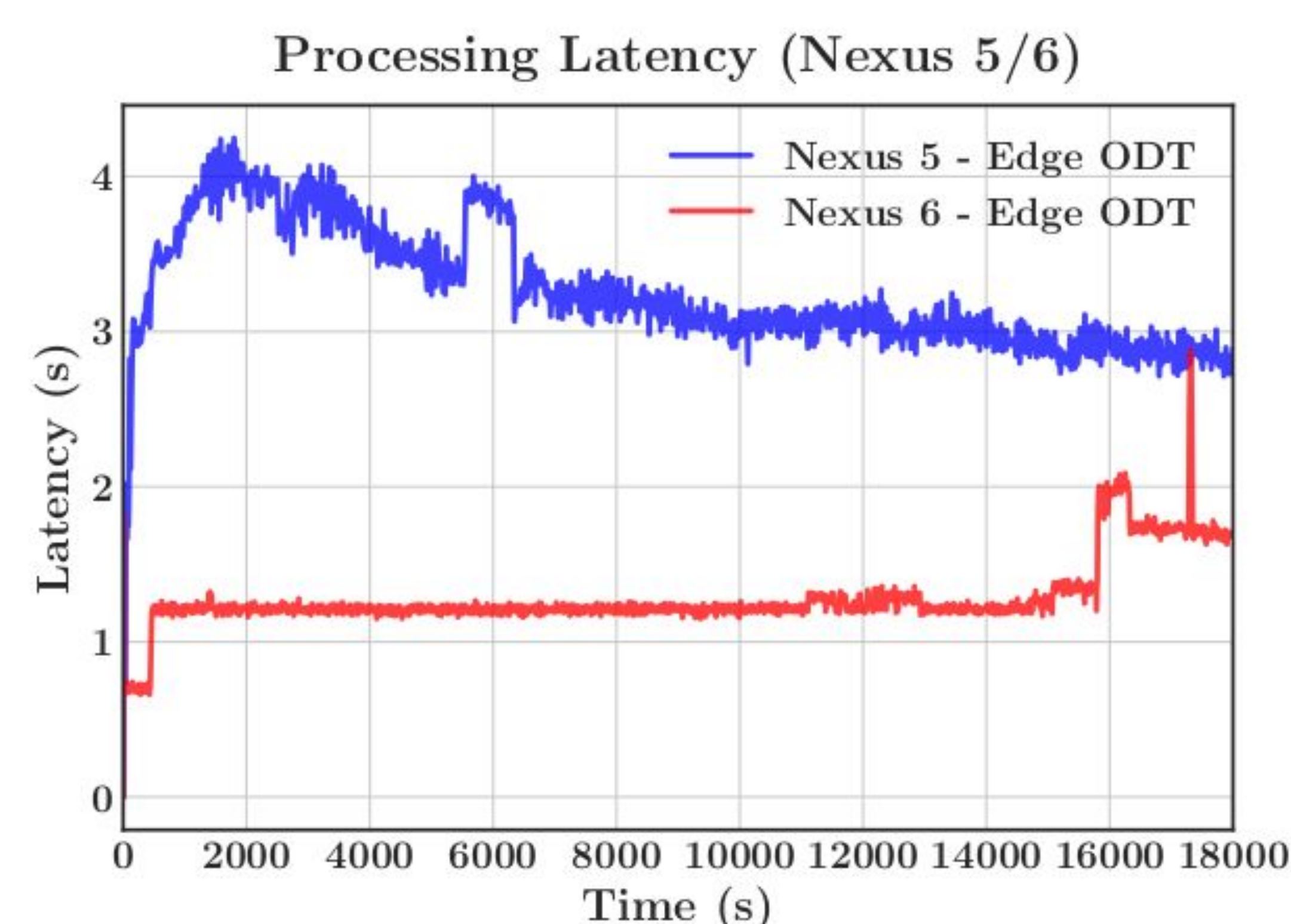
- Build an edge-cloud vision system performing Object Detection and Tracking (ODT) using TensorFlow using SSD-MobileNet-v2 model
- Run it on Google Nexus 5 (Qualcomm Snapdragon 800 2.26 GHz quad-core processor with 2 GB memory) and Google Nexus 6 (Qualcomm Snapdragon 805 2.7 GHz quad-core processor with 3 GB memory)
- All results are shown from data over five hours runs



TEMPERATURE & PROCESSING LATENCY MEASUREMENTS

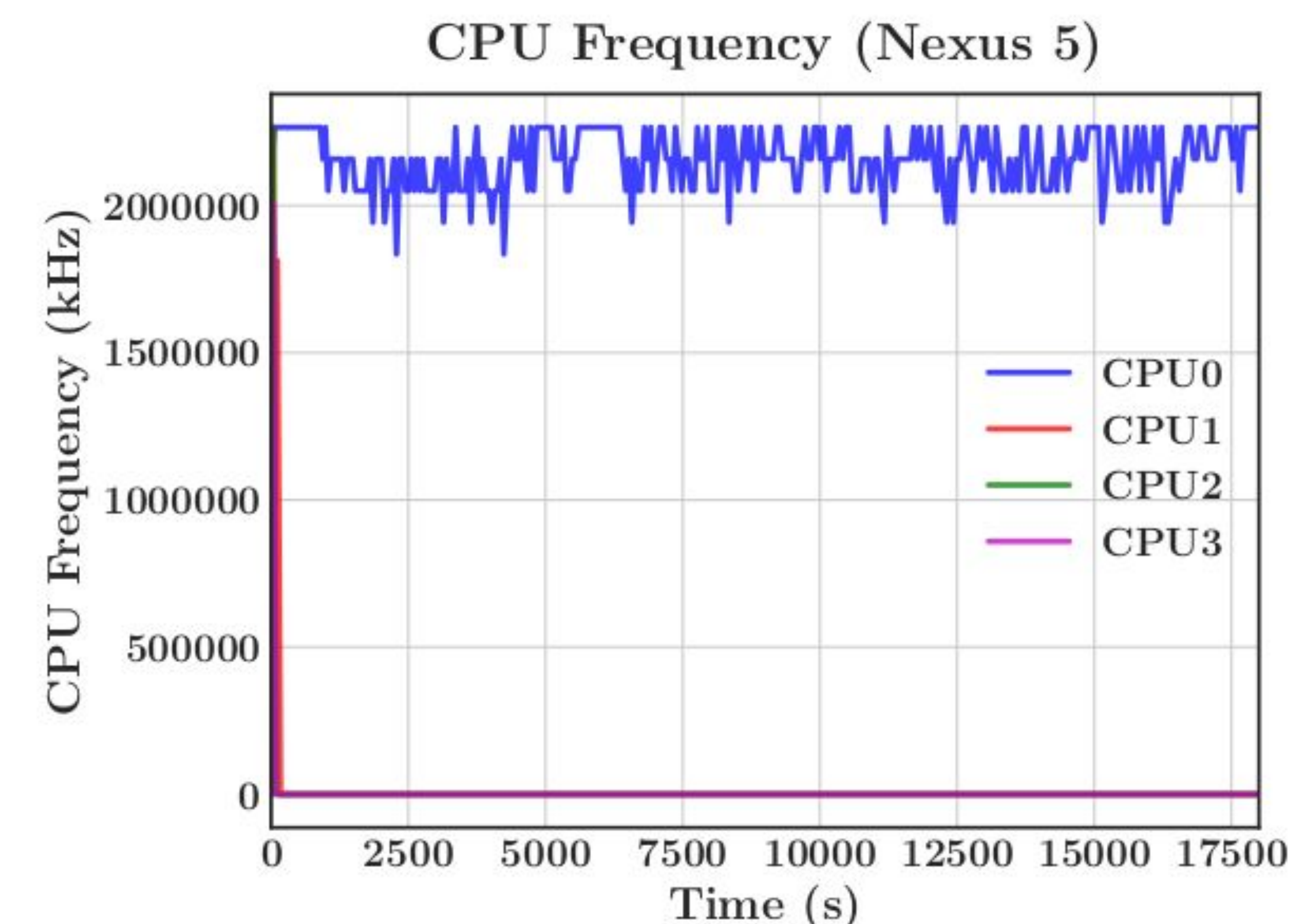


- Due to the processing load, CPU temperature goes above an acceptable threshold (i.e. 60 °C)

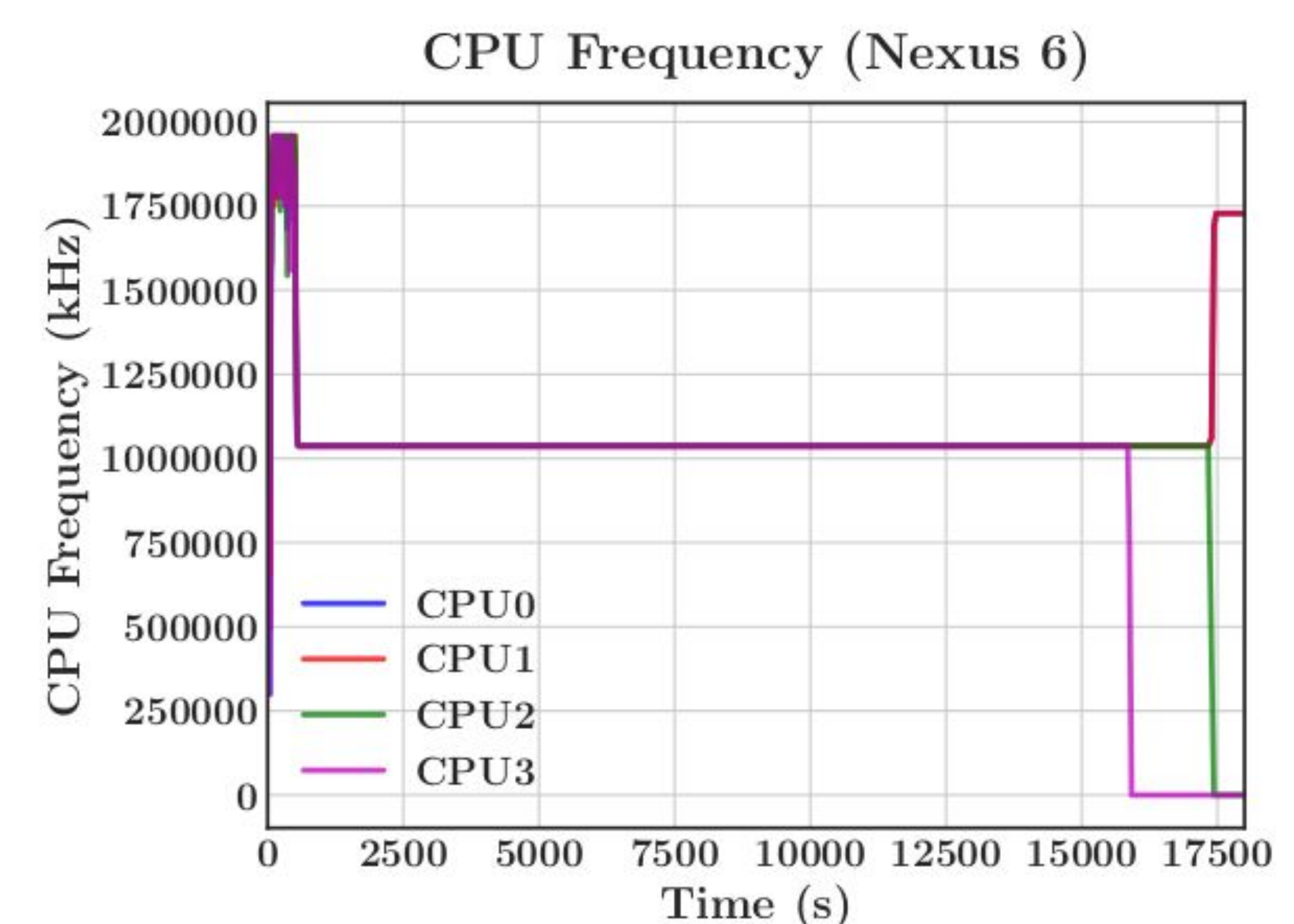


- Temperature variations cause the OS to reduce CPU frequency in both devices, increasing processing latency per frame

CPU FREQUENCY MEASUREMENTS



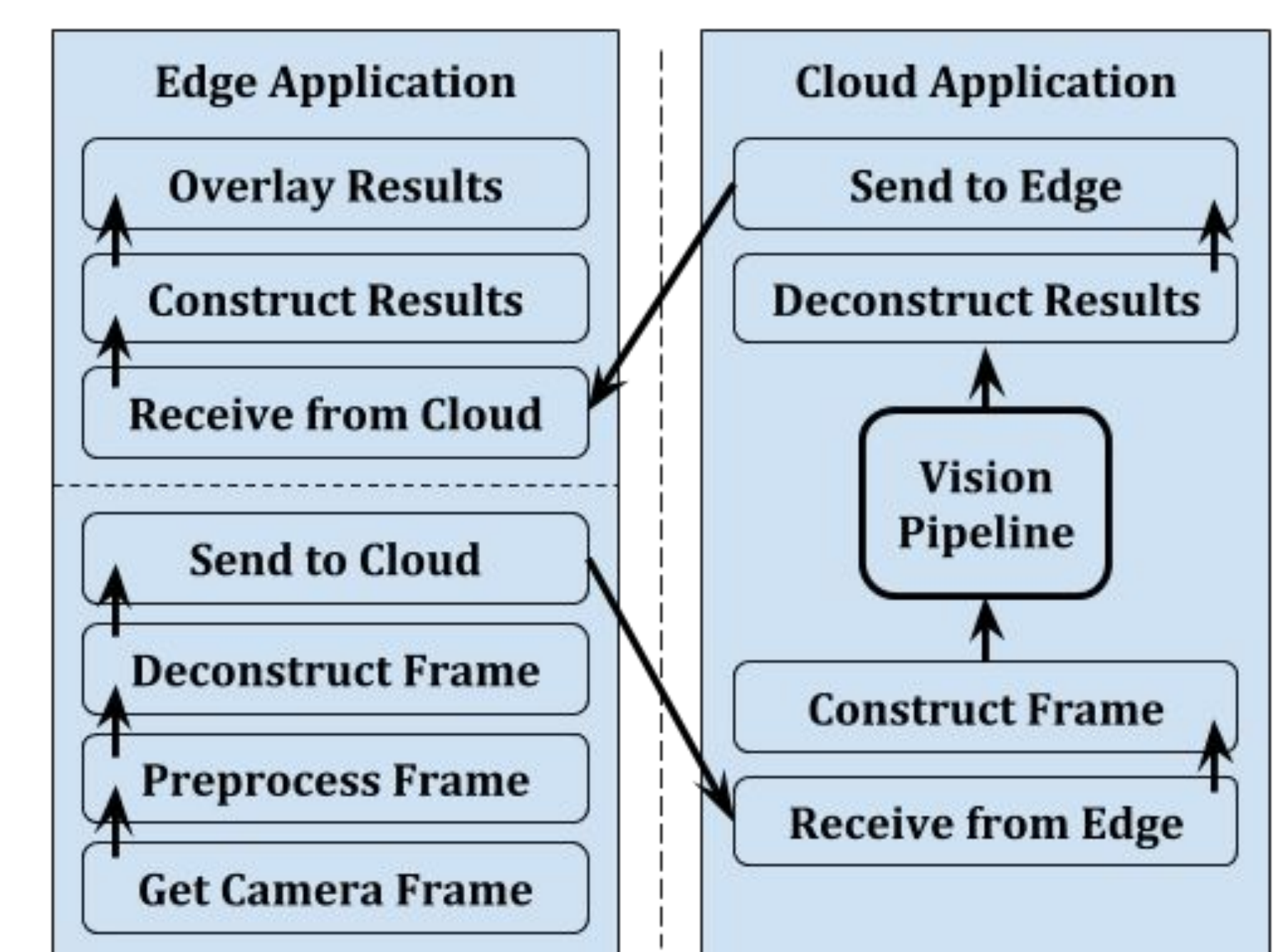
- Nexus 5 OS decides to run on one core for most of the time to reduce platform temperature



- Nexus 6 OS turns off some cores to reduce platform temperature after running for four hours on a reduced CPU frequency

DISCUSSION

- These observations have been made while running one application on the edge
- Running multiple applications would further affect the performance
- Similar behavior observed from running multiple vision applications on the edge, and also running ODT application in a split edge-cloud manner



- **PROPOSAL:** Future edge applications (including ones deployed on edge-cloud) need to explicitly test for platform variation due to long-term operation to demonstrate feasibility

CHALLENGES

- It is still challenging to exclusively run such computationally-intensive applications on the edge
 - **Solution:** deploy applications across edge-cloud
- Recent edge-cloud solutions drawbacks:
 - Tested for a few seconds/minutes and not for long-term operation
 - Ignore platform constraints when multiple applications are running on the same edge device
- Several applications require timeliness guarantees in sensor processing as well as control
- Such applications suffer due to platform variance during long-term operation