

# ML4IoT - HW1 Report

Group 11

November 25, 2024

## Exercise 1: Timeseries Processing for Memory Optimization

### Memory Calculations

#### Primary TimeSeries (Temperature and Humidity):

$$\text{Entries per TimeSeries} = \frac{30 \times 24 \times 60 \times 60}{2} = 1,296,000.$$

$$\text{Compressed Memory (90\% reduction)} = 1,296,000 \times 16 \text{ bytes} \times 10\% \approx 1.98 \text{ MB}.$$

#### Aggregated TimeSeries (Temperature and Humidity):

$$\text{Entries per TimeSeries} = 365 \times 24 = 8,760.$$

$$\text{Compressed Memory (90\% reduction)} = 8,760 \times 16 \text{ bytes} \times 10\% \approx 0.0134 \text{ MB}.$$

#### Total Memory Per Client:

$$\text{Primary TimeSeries (2): } 2 \times 1.98 \text{ MB} = 3.96 \text{ MB}.$$

$$\text{Aggregated TimeSeries (6): } 6 \times 0.0134 \text{ MB} = 0.0804 \text{ MB}.$$

$$\text{Total: } 3.96 \text{ MB} + 0.0804 \text{ MB} \approx 4.04 \text{ MB}.$$

#### Total Memory for 1,000 Clients:

$$1,000 \times 4.04 \text{ MB} = 4,040 \text{ MB} \approx 3.95 \text{ GB}.$$

### Conclusion

The memory required for 1,000 clients is approximately **3.95 GB**, making the system efficient and scalable.

## Exercise 2: Voice Activity Detection Optimization and Deployment

### Hyperparameters and Their Impact

A parameter *overlap* spanning  $[0, 1)$  was introduced such that:

$$frame\_step = (1 - overlap) \times frame\_length,$$

where *frame\_step* and *frame\_length* are the frame step and frame length in seconds, respectively.

The considered hyperparameters are presented in Table 1.

#### Impact on Latency and Accuracy:

Parameter	Values
frame length (s)	{0.016, 0.02, 0.032, 0.05, 0.064}
overlap	{0, 0.25, 0.5, 0.75}
dB threshold (dB)	{5, 10, 15, 20}
non-silence threshold (s)	{0.05, 0.1, 0.2, 0.5}

Table 1: Considered Hyperparameters

- **Latency:** The latency is reduced using frame lengths of powers of 2 in ms, since the STFT algorithm is optimized for power of two input sizes. For the considered framelengths {16, 32, 64} ms, the number of samples per frame is also a power of 2 since the sample rate is 16,000 Hz. In other cases, an increase in frame length increases latency because it increases the number of samples in each frame, given that frame step is constant. Shorter frame steps (greater overlap) increases the number of frames to process, increasing latency. The dB threshold and non-silence threshold parameters do not affect the latency, as they do not affect the STFT computation process.
- **Accuracy:** A shorter frame length and shorter frame step (greater overlap) increase accuracy, as it increases the temporal resolution of the data, so that short sounds are not missed. The dB threshold and non-silence threshold parameters affect accuracy the most.

### Chosen Hyperparameters

From the analysis of hyperparameters impact on latency, it is clear that a choice of a frame length as a power of 2 in ms is fundamental to meet the constraints. The selected hyperparameters, that achieved high accuracy and low latency, are presented in Table 2.

Parameter	Value
frame length (s)	0.032
overlap	0.5
dB threshold (dB)	10
non-silence threshold (s)	0.1
<b>accuracy (%)</b>	97.89
<b>median latency (ms)</b>	19.5

Table 2: Selected Hyperparameters and Metrics

### Conclusion

The system achieved an accuracy of **97.89%** with a median latency of **19.5 ms**, meeting the requirements for real-time Voice Activity Detection.