APPENDIX

Temporal Differential Privacy for Human Activity Recognition

A. Dataset

1) WISDM dataset: The WISDM dataset consists of tri-axial accelerometer data from 36 subjects who performed 6 activities. For the WISDM dataset, we have used 60% of the data for training teacher models, 10% of the data for training student models, and 30% of the data for testing the accuracy of the student model. Subjects 26 to 36 are used for testing.

2) PAMAP2 dataset: This is a multivariate sensor dataset comprising 12 activities recorded from 6 users. For the PAMAP2 dataset, runs 1 and 2 from subject 5 train the student model, resulting in around 1600 examples. Runs 1 and 2 from subject 6 are used for testing the student model. The rest of the data (around 14000 examples) is used to train the ensemble of teacher models.

In both datasets, the testing split of the student model is kept the same as the previous state-of-art baselines [11], [14], [28]. It allows for a fair comparison of the classification metric.

B. Experimental Setup Details

TABLE II
BEST PARAMETERS OF THE LSTM MODEL (POST-ABLATION) USED FOR THE EXPERIMENTS.

Dataset	Layers	Units	Opt	LR	Epoch	Batch
WISDM	2	64	Adam	0.0035	500	64
PAMAP2	2	64	RMSProp	0.01	150	64

TABLE III
SEQUENCE LENGTH RANGES FOR BEST CLASSIFICATION PERFORMANCE FOR THE DATASETS (CHOSEN FROM PREVIOUS WORKS [11], [28].

Dataset	Time Range (in seconds)	Window Sizes
WISDM	6 to 10	100 to 200
PAMAP2	5 to 9	180 to 240

1) Partitioning strategy using window-sizes: To train *TEMPDIFF* we select 5 window-sizes and The window sizes used for temporal representation in *TEMPDIFF* (represented by $w_1, w_2, \ldots, w_N$ in Figure 3) are chosen based on classification performance of previous state-of-the-art ([11], [28]). The time range within which the N window sizes are selected is shown in Appendix B Table III. In *Vanilla-PATE*, the dataset is partitioned into a fixed number of subsets as proposed by Papernot et al. in [22], which is equal to the number of teacher models. E.g., a hundred partitions, $M = 100$, means a hundred models. On the other hand, in *TEMPDIFF*, we can have lesser partitions in the data to have the same number of teacher models. For example, we can train 100 models using 20 partitions and 5 different window sizes, i.e., $M = 20$ and $N = 5$. Other combinations of M and N that multiply up to the total number of teacher models are possible if they are within acceptable window sizes (shown in Appendix B Table III). After randomly choosing N window sizes from a range of acceptable window sizes, we calculate the number of initial partitions, M. The total number of desired teacher models divided by the number of window sizes, N, is a fair estimator for the number of partitions, i.e., M in *TEMPDIFF*. We set $N = 5$ in our experiments and estimate M accordingly.

C. Algorithms

Algorithm 1 Temporal Partitioning Algorithm

A time-series dataset D
D divided into M disjoint partitions $Partition_1, Partition_2, \ldots, Partition_M$
N window sizes, $w_1, w_2, \ldots, w_N$
An empty dataset list DL of size *M x N*.
for $i = 1$ to M **do**
    Divide $Partition_i$ into N sub-partitions $Partition_{i1}, Partition_{i2}, \ldots, Partition_{iN}$
    for $j = 1$ to N **do**
        Slide w_j over $Partition_{ij}$ to extract a row.
        Continue sliding to form $Dataset_{ij}$
        Add $Dataset_{ij}$ to DL.
    end for
end for
Each dataset in DL forms a temporal partition.

D. Implementation Details

In our implementation, we have used the Python language with the Pytorch library [24] for building our models. Additionally, we have used the PySyft library from https://www.openmined.org/ to calculate our privacy budgets.