Figure 1: Mean task timing metrics in each level of a 4-level tree aggregation on 16 nodes with 256 partitions of the underlying RDD. The operation is a distributed gradient vector computation for 20 million instances in the labelled point RDD, and the problem has a dimension of approximately 30 million. The mean total task time is τ, and the execution, garbage collection, serialization/de-serialization, and communication times are tex, tgc, $tser$, and tio, respectively.