The latest benchmarking experiment – quick summary

Gyuri Barabás

The latest benchmarking was run under the same settings as the one that was run before in April. The one difference is that uploading files of 100MB or larger was not possible.[1] So instead of 100MB files, we had 50MB ones in this run of the experiment.

We can visually compare download times across the April (2025-04) and June (2025-06, that just finished yesterday) runs. We do this both for the `DATA` (Figure 1) and `RACE` (Figure 2) retrieval strategies.

The results from the two runs look quite similar, though we appear to see consistently faster download times under the `RACE` retrieval strategy for the June experiment. However, it is unlikely that this observed difference is significant. Indeed, if we make a group-by-group comparison of the two experiments (taking Figure 1 and Figure 2, and for each file size category / retrieval strategy / erasure coding combination, compare the April and June data using a non-parametric Wilcoxon rank sum test), none of the results come out as significant at the $\alpha = 0.05$ level, after correcting for multiple comparisons (Figure 3).

We can also check how much advantage the `RACE` strategy offers over `DATA`, for each file size category. The results are qualitatively similar for the April (Figure 4) as for the June (Figure 5) runs. The one difference is in the largest file size categories, which are difficult to compare because in April that size was 100 MB and in June 50 MB. That said, unless there is a sudden jump in behavior somewhere between those two file sizes, the June results look much better than the April ones did.

When looking at upload speeds, they are nearly the same across the April and June runs, but June is slightly faster (Figure 6).

Since the two experiments were run under the same release version, the differences between them are caused purely by the network, and its number and distribution of nodes.

[1] Information from Marko Zidaric.

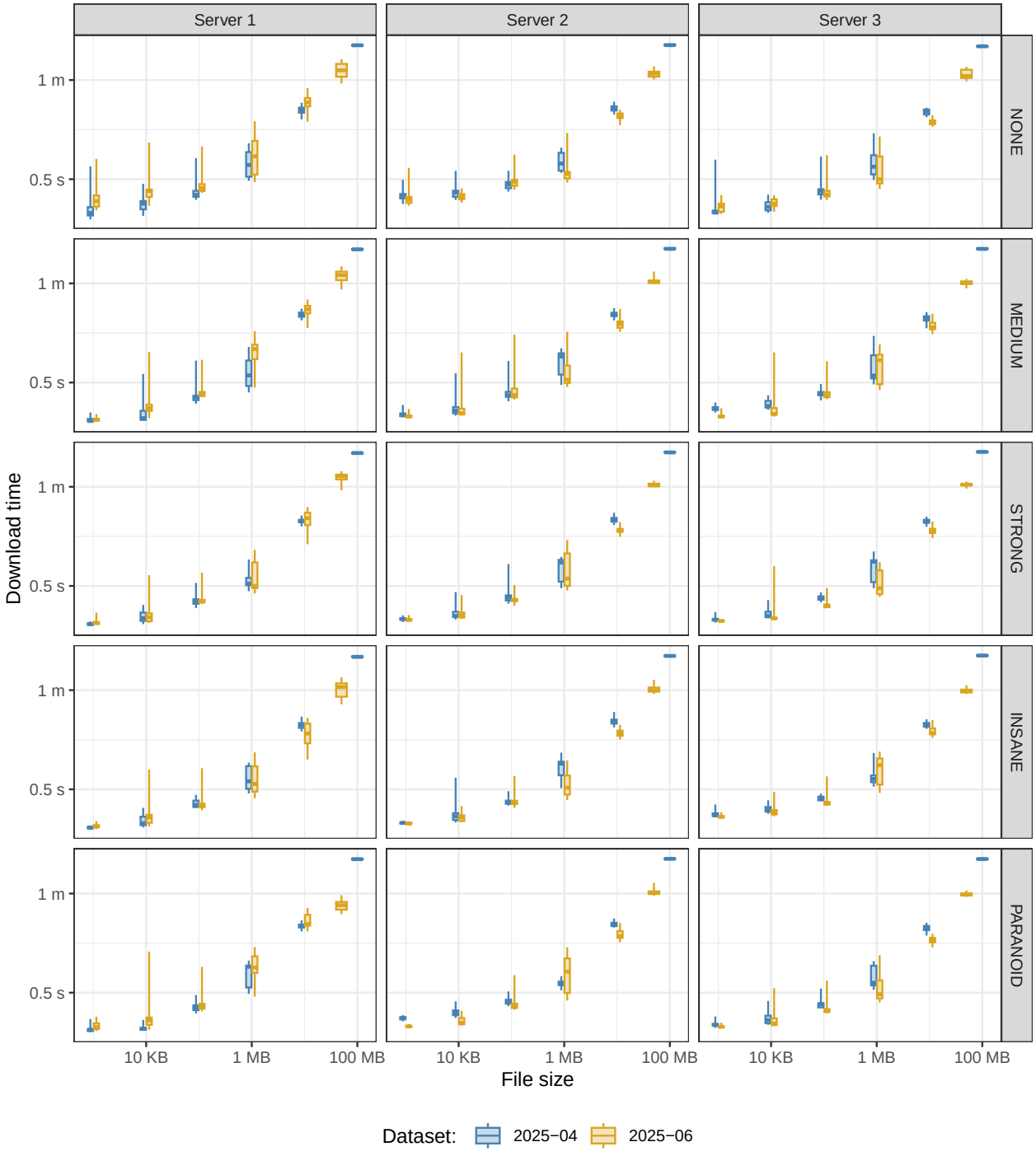

Figure 1: Comparison of the April (blue, 2025-04) and June (yellow, 2025-06) benchmarking experiments, for the `DATA` retrieval strategy. Box plots are standard except no outliers are shown—that is, the thick horizontal line is the median (point that separates the top and bottom half of the data), the box around it encompasses the middle 50% of all data points, and the top/bottom whiskers show where the top/bottom 25% of the data are.

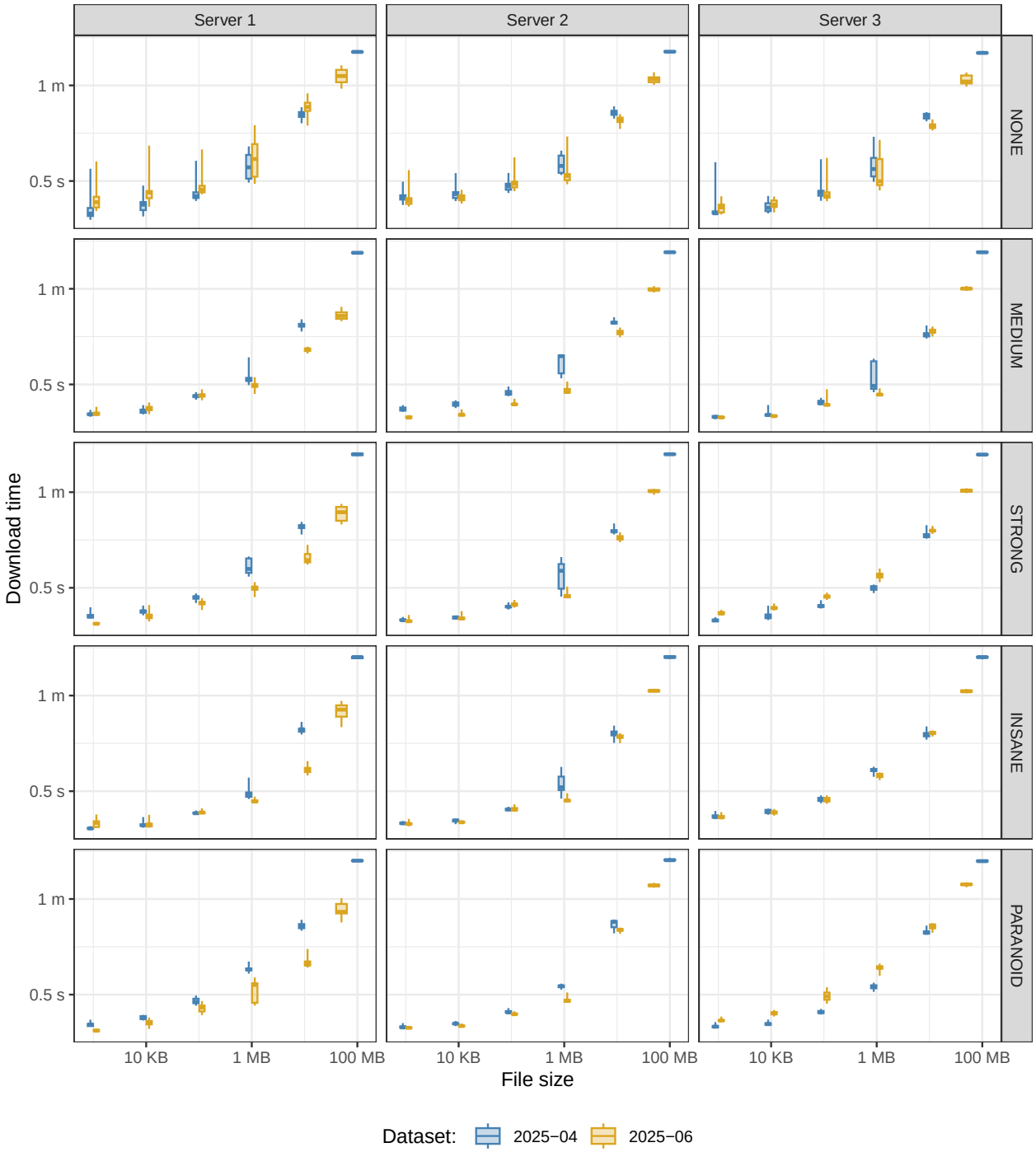

Figure 2: As Figure 1, but for the `RACE` retrieval strategy.

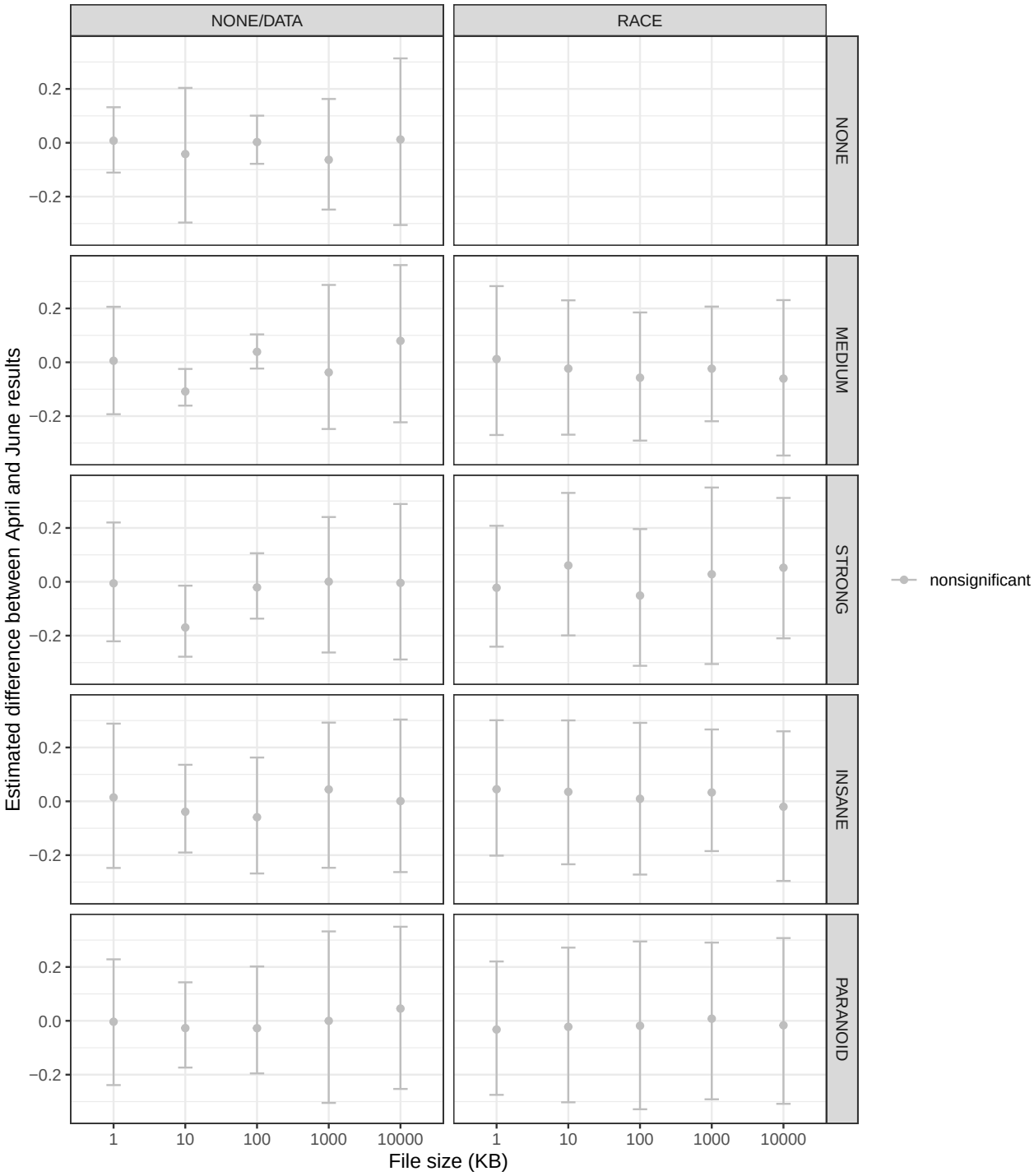

Figure 3: Group-by-group comparison of results from the April (2025-04) and June (2025-06) benchmarking experiments. The y-axis shows the estimated difference (point) plus/minus 95% confidence intervals (error bars) from a Wilcoxon rank sum test applied to each distinct file size / retrieval strategy / erasure coding combination. Colors indicate whether the difference was found significant at the $\alpha = 0.05$ level, after false discovery rate correction to multiple testing. Since no results turned out as significant, everything is gray and no results are in blue.

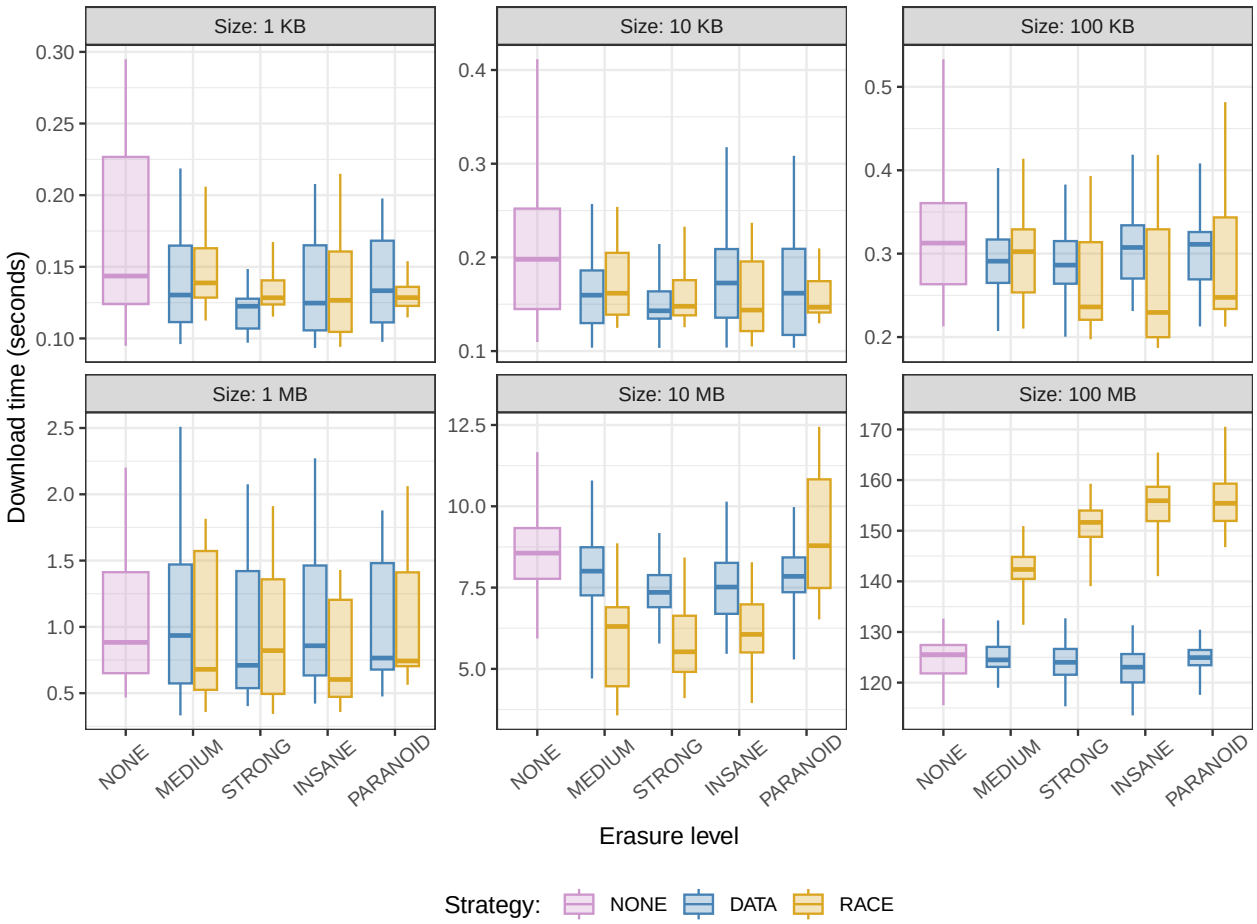

Figure 4: Download times (y-axis) as a function of erasure level (x-axis), file size (panels), and retrieval strategy (colors). Data are only for the April (2025-04) experiment. The y-axis is individually scaled for each panel. High outliers (points more than 1.5 times the interquartile range outside the box in the upper direction) have been removed, because they otherwise distort the plots and make the corresponding results difficult to compare visually.

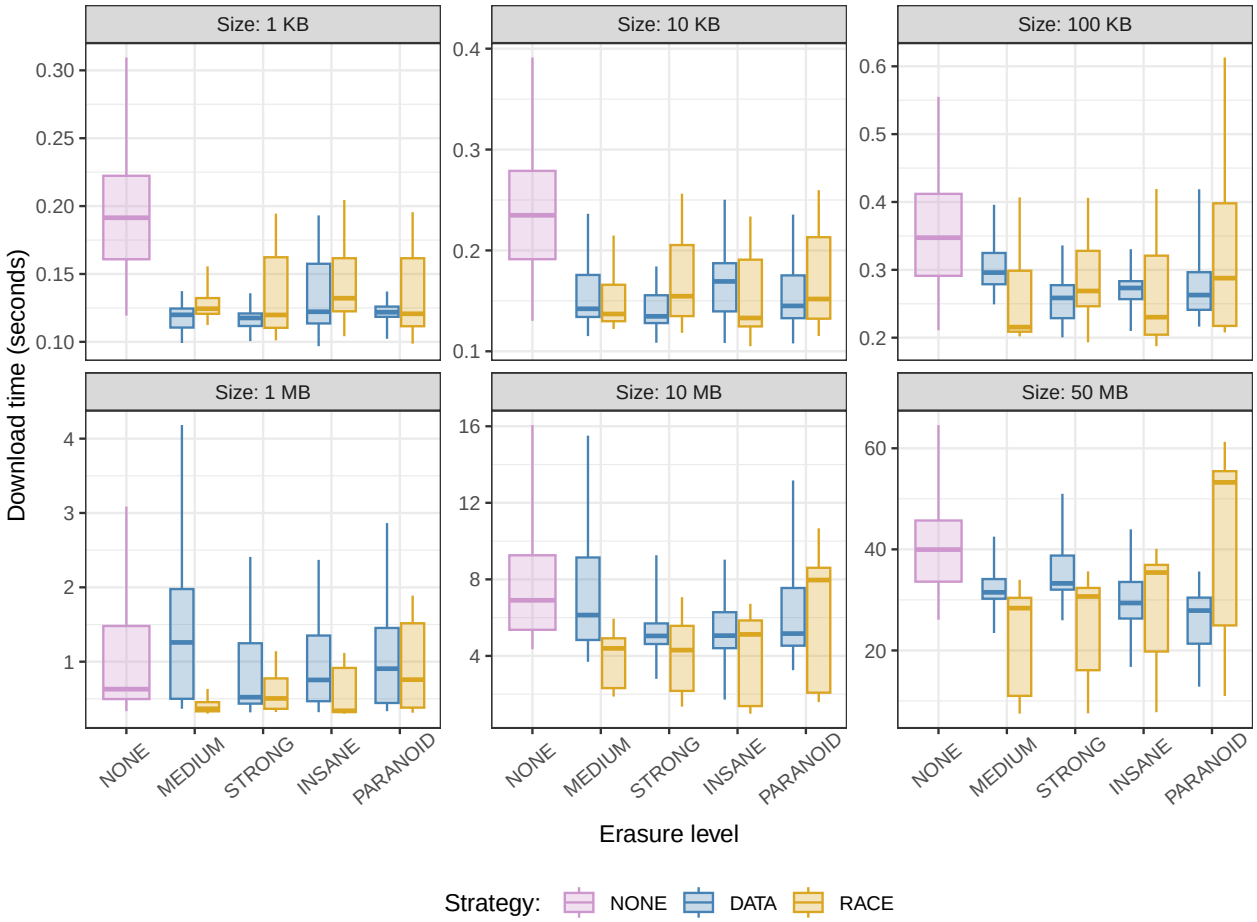

Figure 5: As Figure 4, but for the June (2025-06) experiment.

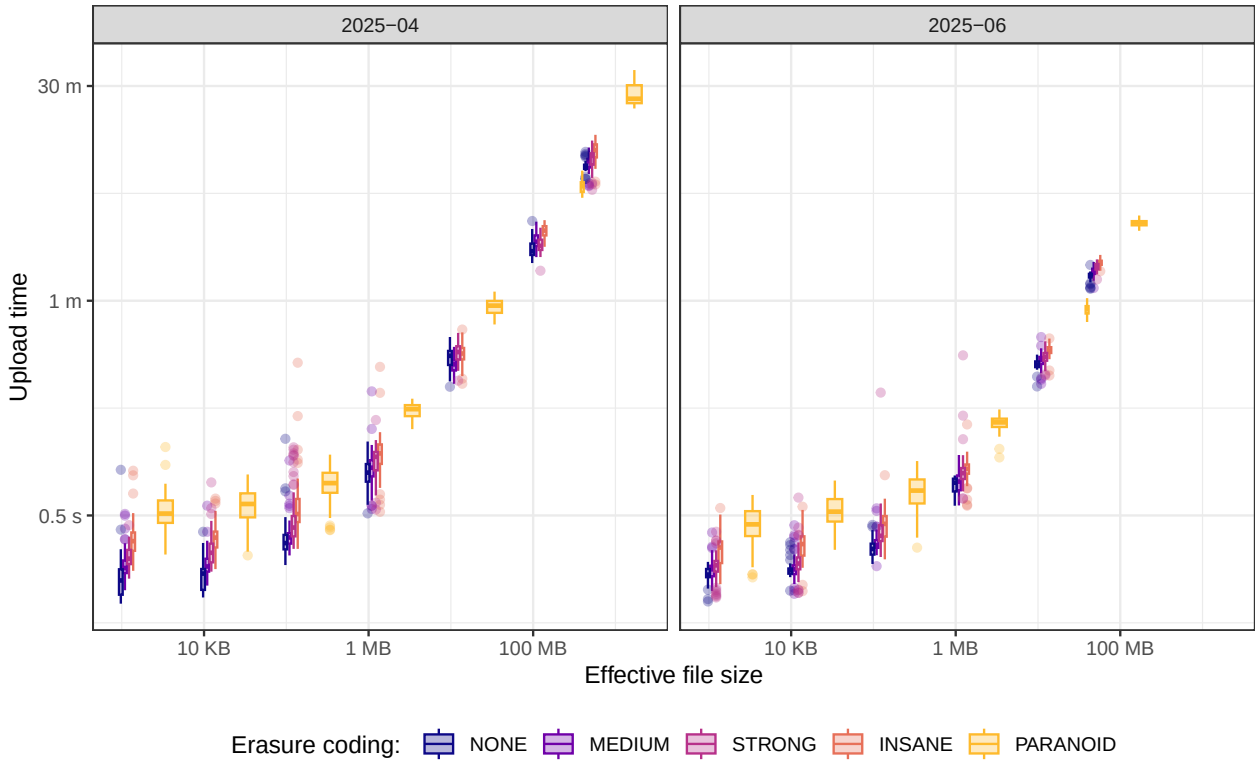

Figure 6: Upload times in the April (left panel) and June (right panel) benchmarking experiments. The graphs take into account the effective sizes of the files—that is, overhead from erasure coding and packed-address chunks have been accounted for.