

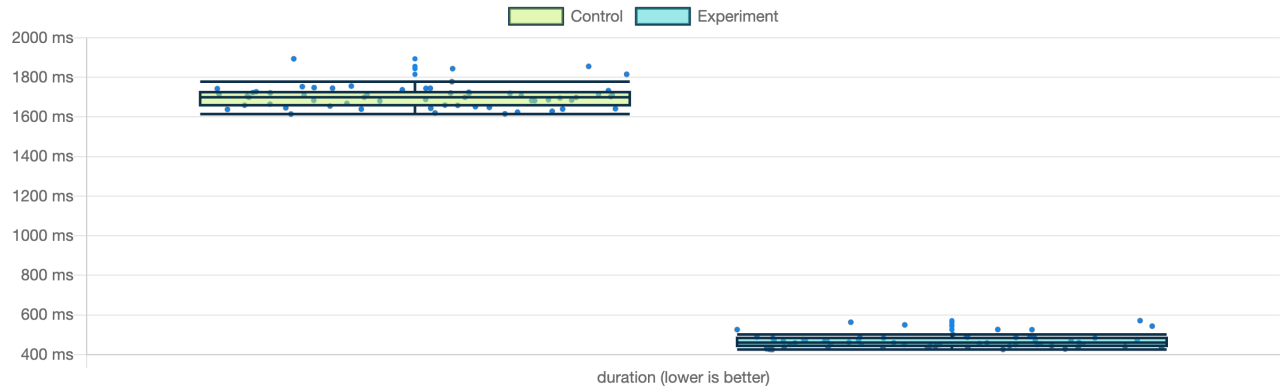
TracerBench: Boxplot Results

TracerBench on HeadlessChrome/84.0.4112.0



duration (1232 ms faster)

A statistical analysis test was used ([Wilcoxon Rank-Sum Test](#)) to determine that the results are significant meaning they are worth looking at. A statistics estimator ([Hodges-Lehmann estimator](#)) was used to determine "Experiment" is **faster** by **1232 ms**. TracerBench is 95% confident it is **faster** between **1215 ms to 1248 ms** based on 60 samples using a [confidence interval](#).



Sub Phases of Duration

The chart below shows the finish times (a point in the page load duration) of the sub phases for experiment and control. It gives a high level view on what also changed (if there are any).

You can view more details about the phases starting from [page 2](#).

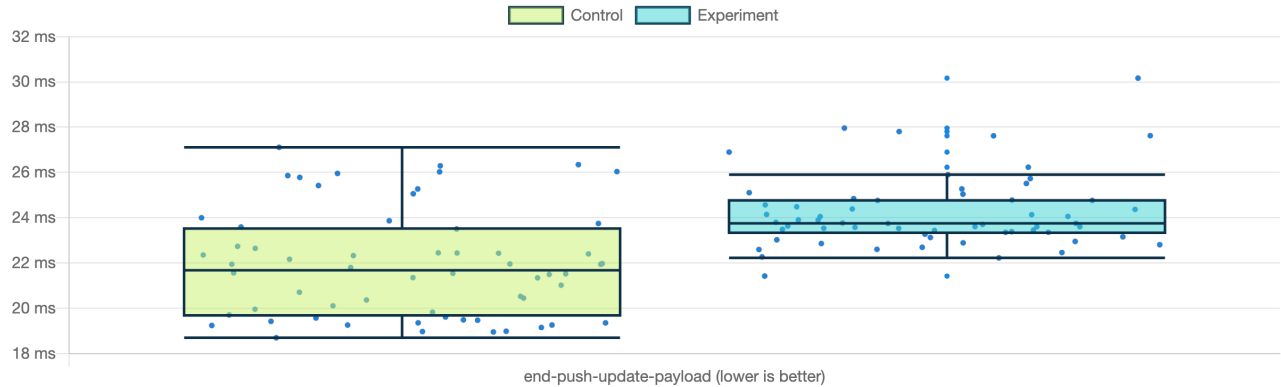


Sub Phases in Detail

Sorted by phases that regressed with the most magnitude.

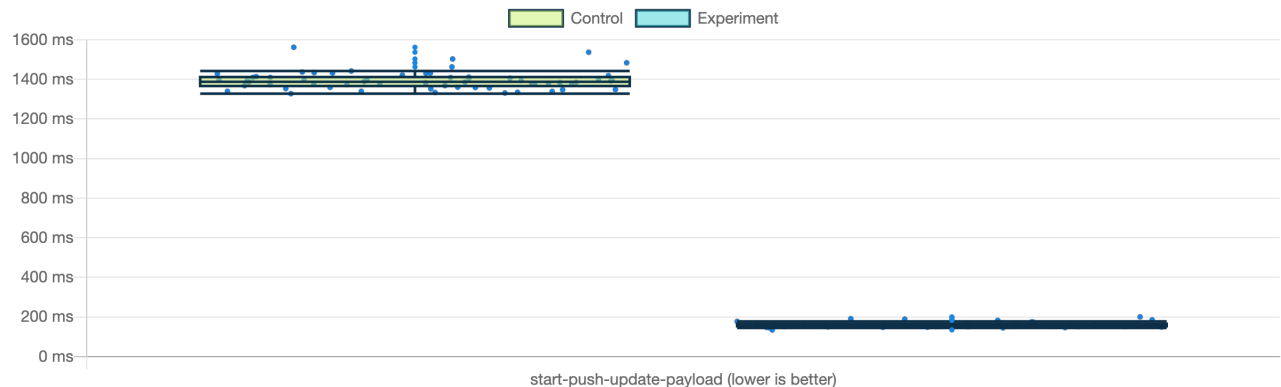
end-push-update-payload (2 ms slower)

A statistical analysis test was used ([Wilcoxon Rank-Sum Test](#)) to determine that the results are significant meaning they are worth looking at. A statistics estimator ([Hodges-Lehmann estimator](#)) was used to determine "Experiment" is **slower** by **2 ms**. TracerBench is 95% confident it is **slower** between **2 ms to 3 ms** based on 60 samples using a [confidence interval](#).



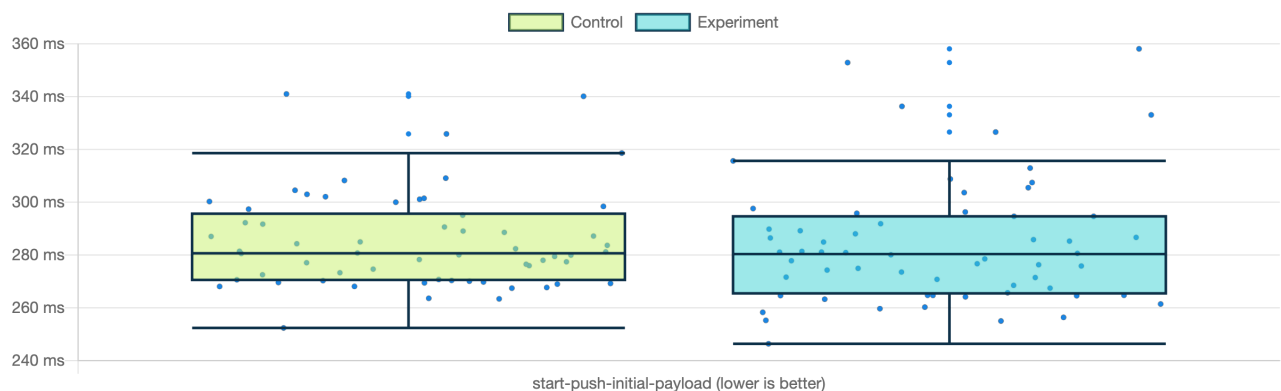
start-push-update-payload (1229 ms faster)

A statistical analysis test was used ([Wilcoxon Rank-Sum Test](#)) to determine that the results are significant meaning they are worth looking at. A statistics estimator ([Hodges-Lehmann estimator](#)) was used to determine "Experiment" is **faster** by **1229 ms**. TracerBench is 95% confident it is **faster** between **1220 ms to 1238 ms** based on 60 samples using a [confidence interval](#).



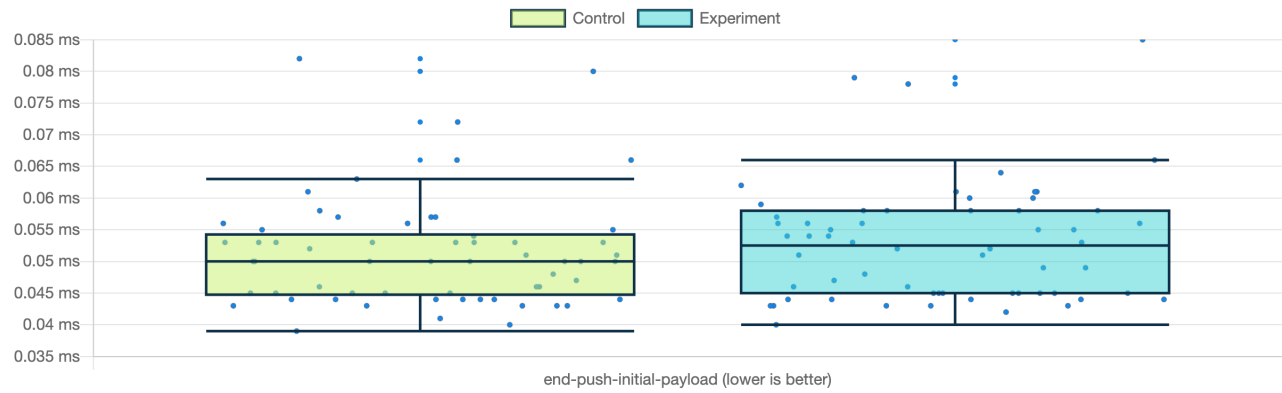
start-push-initial-payload (No Difference)

A statistical analysis test was used ([Wilcoxon Rank-Sum Test](#)) to determine that results are not significant.



end-push-initial-payload (No Difference)

A statistical analysis test was used ([Wilcoxon Rank-Sum Test](#)) to determine that results are not significant.



Resources

- [Stats Primer](#)
- [Understanding Boxplots](#)
- [Wilcoxon Rank-Sum Test](#)

Configs Used

```
{  
  "tbResultsFolder": "/Users/amcgs/Sources/pieter-v/data/packages/unpublished-relationship-performance-test-app/tracerbench-results"  
}
```