

# 1 Bug 798130 / 799227 – Pinpoint startup regressions in Firefox 18

## 1.1 tl;dr

Analysis of SIMPLE\_MEASURES\_FIRSTPAINT from Telemetry does not show evidence of performance regression in FF18 nightly builds, including on or around 2012.09.11. This does not rule out a regression, but there is not conclusive evidence in this data, and the data examined was not adequate to isolate a regression in any particular build.

## 1.2 Details

I examined SIMPLE\_MEASURES\_FIRSTPAINT data subject to the following filters:

- submission dates: from 2012.08.20 to 2012.10.12,
- app name: “Firefox”,
- OSes: WINNT, Linux, and Darwin,
- app update channel: nightly,
- app version string containing “17” or “18”.

This sample included appBuildIDs from 2012.07.17 to 2012.10.07. These appBuildIDs were clubbed by day; I call this “buildDate” (all appBuildIDs whose first 8 characters matched were grouped– i.e., “2012071712345” would be grouped with “2012071798765”).

For each buildDate, the graphs in Figure 1 show:

1. the quantiles of SIMPLE\_MEASURES\_FIRSTPAINT time in seconds (the top frame show the 99th percentile, the next frame show the 1st, 25th, median, and 75th percentiles) for each buildDate,
2. the 3rd frame shows the number of submissions receive for each buildDate through out the submission period (the y-axis scale is in THOUSAND of submissions)
3. the bottom frame shows the mean of SIMPLE\_MEASURES\_FIRSTPAINT in seconds for each buildDate.

One possible explanation of the signal that you observed in the Telemetry data could be that the data was aggregated across buildDates in the telemetry dashboards. When all the measurements of SIMPLE\_MEASURES\_FIRSTPAINT received in a day for given build series (17 or 18 nightly) are plotted, there does appear to be a slight regression for the 18 nightly series as compared to the 17 nightly series. See Fig 2.

I have not analyzed the TP5 data that you gave me a link to, so I cannot comment on the quality of any apparent signal in that data.

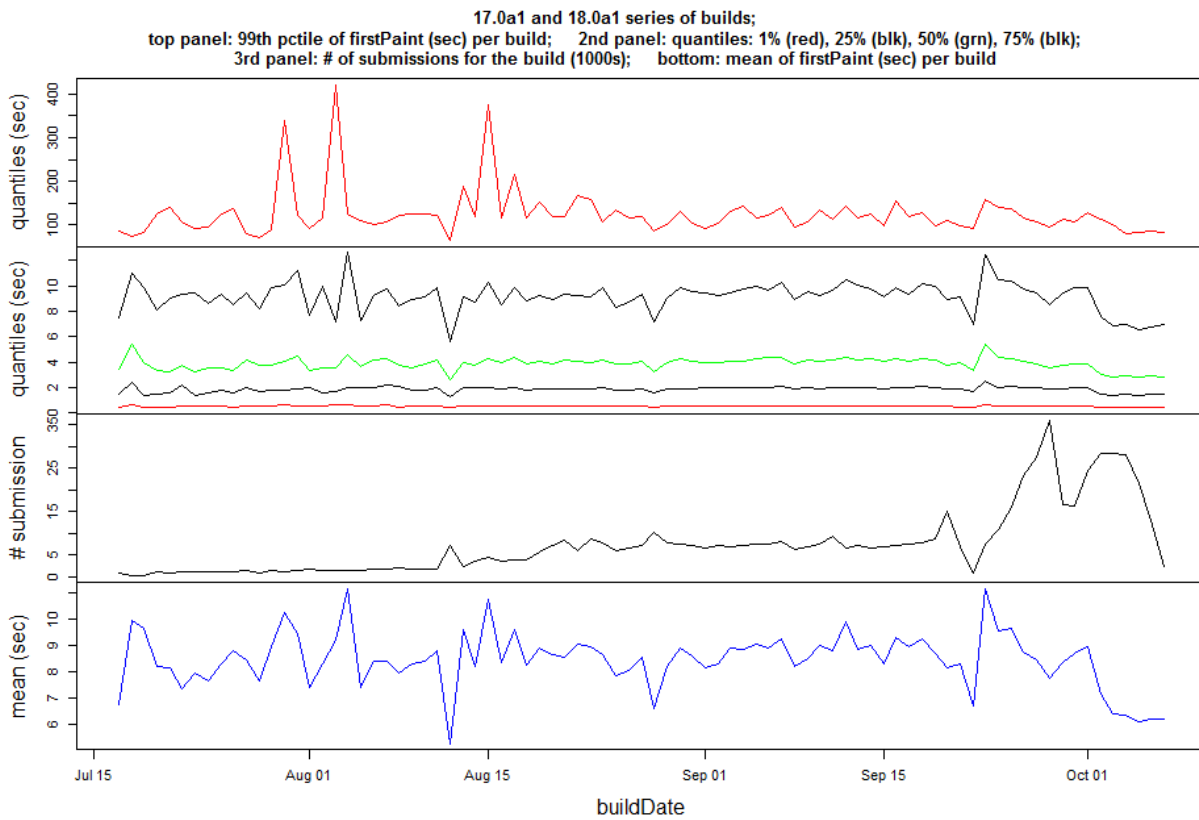


Figure 1: The 18 nightly series of builds starts on August 27. There is no visible regression in startup time in either the tail of the distribution, the median, or the mean.

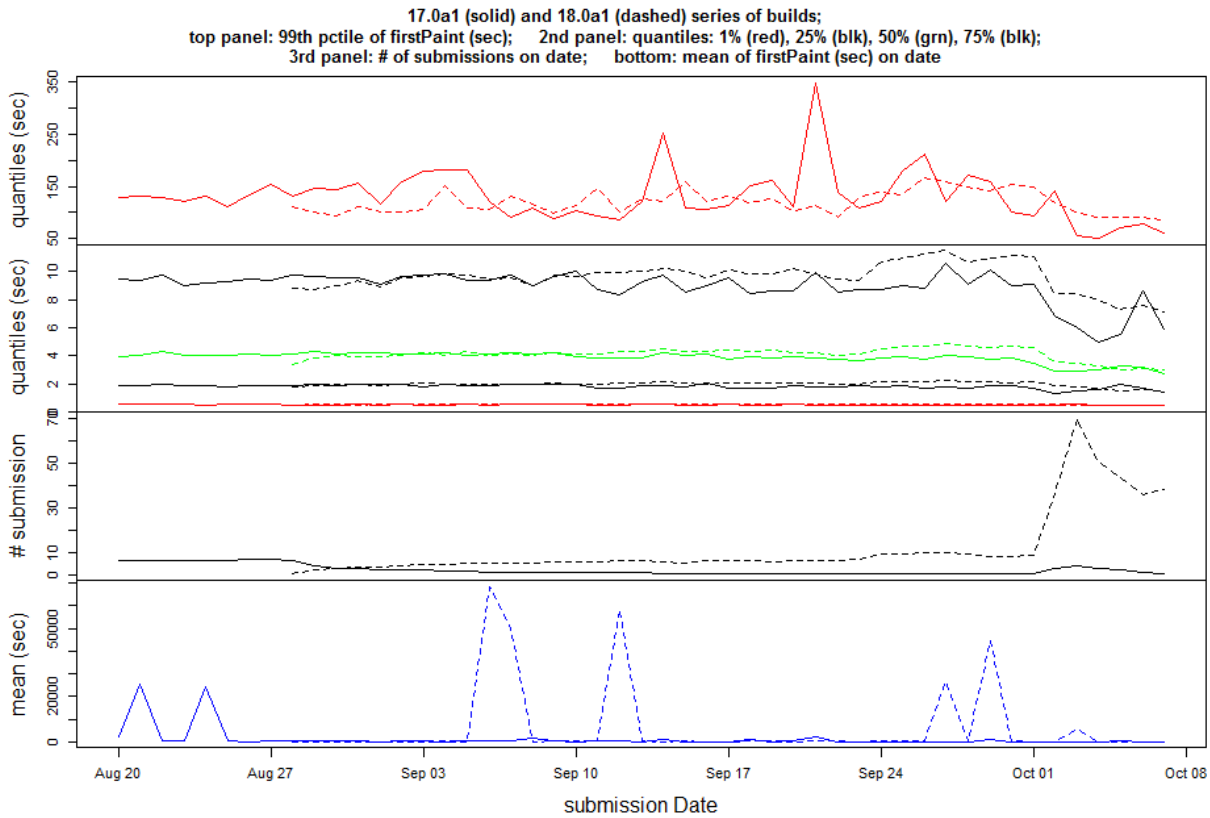


Figure 2: This set of plots depicts all of SIMPLE\_MEASURES\_FIRSTPAINT received on a day for \*all\* of the buildDates currently existing– for example, the values shown on September 11 include all of the measurements recorded for 17 Nightly with buildDates 2012.07.17-2012.09.11 (solid line) and all of the measurements recorded for 18 Nightly with buildDates 2012.08.27-2012.09.11 (dashed line). There is a slight a perceptible degradation around September 11 in the 75th percentile. I cannot comment on whether this is statistically significant or not; in any case, it is the wrong unit of analysis for isolating a performance regression, but it points out how slicing the data another way could cause the appearance of a regression.