Author's Reply

February 10, 2017

Overview

I appreciate this detailed review and included suggestions.

Reviewer

<< Reviewer's comments to the author(s) >>

The SEDONA is based on the probability that any daemons run during the measurement. And cutoff time is based on the actual measurement time. The proposed approach is interesting, but I cannot understand why the author employs such statistics and heuristic way, and how efficient the proposed approach will work. Please refer to my comments, and consider to revise the manuscript and re-submit .

Thank you for your comments. Indeed, the used statistics and heuristics greatly helped me identify what daemons were infrequent, long-running on our system (as described in the third paragraph in the left column on page 3). Applying such statistics and heuristics can be done to any Linux system, which indeed makes our protocol generic. Furthermore, the experimental results show that the proposed protocol is more efficient than the traditional method relying on elapsed time.

I now respond to individually below. Note that my responses to your comments are incorporated into the revision and marked in blue.

Comments: (1) In the last of the manuscript, the author describes "Our plan is to integrate SEDONA into the query timing protocol [8]." If the research goal is the combination of the SEDONA and the query timing protocol, the proposed approach may be suitable. However, the goal is not clearly described in the manuscript. The readers will recognize the SEDONA is a generic protocol. If the research goal is the combination of this work and the prior work, the author should mention so.

The previously-described plan was suggested as one of the possibilities in the future work. But that was not stated as the ultimate goal of this paper. As you already recognized, my goal is to propose a *generic, better* timing protocol against the conventional approach of utilizing elapsed time, to enable better timing results. This goal is successfully achieved by significantly reducing variability in timing as demonstrated in our various experiments. I now clearly state this goal in the last sentence in Section 1.

To avoid confusion, I removed the last mention. Also, the first paragraph of the previous manuscript on page 2, which used to provide an explanation of a possible combination of the query timing protocol with SEDONA, was removed.

(2) If the proposed approach is for a generic purpose, there is a strong doubt about usability. The improvement of the SEDONA is little. And the extent of the each benchmarks improvement described in Table 3 varies widely. Thus, the experimental results does not confirm whether the SEDONA is widely practical or is useful. The description to clarify the SEDONAs characteristics is needed.

Such a doubt may be raised. But I do not agree with the comment that the improvement is little, for the following reasons. First, for a certain workload such as `434`, our protocol reduced the variability up to by about 10x (from 75 ms to 7 ms), which cannot be said to be small. (Please refer to the last sentence in the fourth paragraph in Section 3 on page 4.) There were also some other workloads such as `410` and `445` in which the improvement reached about 3x, respectively. Furthermore, every benchmark's standard deviation by our protocol was less than that of the extant method. The relative error of the benchmarks was on average improved by 1.5x as well.

Second, our protocol scaled well over increasing execution time. (Please refer to the fifth paragraph in Section 3 on page 4.) The relative error of our protocol was lower than that of the traditional method for a long-running benchmark

such as `416`, `436`, and `454` as well as for a medium-length or a short-length one. A new experiment, which I describe shortly, also revealed such scalability of our protocol, as shown in Fig. 5.

(Note that the ET results by the traditional measurement were already improved enough, as they were measured based on the settings handling several identified timing factors mentioned in the third paragraph in Section 1. Under such a well-configured timing environment, our protocol could further reduce the variability of execution time in such various benchmarks. This actually demonstrates its general purpose and wide usability.)

Lastly, the new experimental results provide another strong evidence that our protocol is widely practical and useful. In the new experiment we evaluated the performance of our protocol on some real-world programs like insertion sort and matrix multiplication (in column major). The results showed that the performance gap between the original and our timing schemes was increasingly widened, e.g., up to by 6x, as workload level increased.

Therefore, I claim that these various experiments confirm the wide acceptance and usability of SEDONA as an alternative to the ET-based extant measurement technique.

(3) The improvement of the SEDONA is little. And the extent of the each benchmarks improvement described in Table 3 varies widely. Thus, the experimental results does not confirm whether the SEDONA is widely practical or is useful.

This comment is basically the same as (2). Please refer to our response of (2) for this comment. One note, though. You may feel that the extent of the improvement varies widely. That is because the SPEC CPU2006 includes a set of very different compute-bound benchmarks. We also found that all the different benchmarks in the suite benefited from our timing protocol, and for a certain workload, our protocol significantly reduced their variability up to by 10x as mentioned earlier.

(4) Since the explanation of "dual-PUT" is not enough, I can not well understand the algorithm shown in Fig. 3. Similarly, what is 1st Halfs ET (ms) attached at Fig. 4(b), 4(c), 4(d) ? I understood only that the label is related with dual-PUT. These words are important in the manuscript. So, please keep in mind to write understandably.

I understand your confusion. More descriptions of the term "dual-PUT" are now provided with a simple example in the second paragraph in the right column on page 2. In brief, the term of "dual-PUT" represents a pair of two consecutive samples of a run of a (regular) PUT. (The task length of that regular PUT is virtually doubled, extended to the corresponding dual-PUT.)

In addition, the 1st half indicates the first element of the pair whereas the 2nd half the second element of the same pair. The label of "1st (2nd) Half's ET (ms)" is renamed as "Odd (Even) Samples' ET(ms)," for better clarification. Please refer to the subsequent (third) paragraph.