

Consolidated Review of *Benchmarking Personal Cloud Storage*

1. Strengths:

The paper presents a succinct and interesting comparative study of one type of Internet service that is becoming increasingly popular—personal storage. The topic is timely, given that personal cloud systems are widely used on the Internet. The active measurement approach reveals various important design choices of these systems such as the location of data centers, chunking, bundling, and encoding. The evaluation not only includes end-to-end performance comparisons, but also the comparisons of designs and data transfer optimizations.

2. Weaknesses:

The comparison presented in this paper can be improved to be thorough. At the moment, it is conducted from a single vantage point (i.e., in a university in Europe) and over a short time period. This makes it unclear how the results might change when evaluated from different locations in the world.

The detailed setup can be described better. The exact experiments conducted are not clear at different places. Furthermore, there is no discussion of how the benchmarks selected or whether there exist other benchmarks that could have been considered.

The comparison does not consider various other aspects of the system such as usability, e.g., how easy is it to use, how many clicks are involved for each operation. Finally, the contributions are incremental compared to prior work by the authors.

3. Comments

All the reviewers were fond of the topic, problem statement, and the authors' approach. This paper was highly evaluated, but was also recommended for revisions of the following points.

First, we suggest that the authors justify why they did not conduct experiments over a geographically distributed testbed like PlanetLab or through machines at different locations to see how the results might change when evaluated from different locations. At the very least, the authors should explicitly acknowledge this limitation in Section 2.4.

Second, we suggest that the authors specify the benchmarks in more detail, as this is the title of the paper. In particular, a discussion of alternate benchmarks would be useful.

Third, given these limitations of the experiment, we recommend that the authors take a neutral tone of voice in describing results.

Finally, we encourage the authors to discuss the following questions. Why is it surprising that some systems open up separate TCP connection for each file? Why is it smarter to reuse TCP connection? Maybe there could be other benefits that could come from having separate TCP connections or maybe the design choice doesn't allow certain systems to reuse TCP connections?

4. Summary from PC Discussion

There was no PC discussion.

5. Authors' Response

We would like to sincerely thank the reviewers and our shepherd (Meeyoung Cha) for all their valuable feedback, which helped us to improve the camera-ready paper.

The major concern raised by the reviewers was that we have collected our measurements from a single vantage point over a short time period. The reviewers commented that we should acknowledge this limitation clearly. We completely agree with the reviewers that measuring from other places and longer time intervals would enrich the paper. We plan to repeat our experiments from different locations in future works, and this is now explicitly written in both Sect. 2 and Sect. 6.

The reviewers also commented that it was not clear how our results might change when measuring from other locations. We now clarify in Sect. 2 that our main conclusions would not be affected by that. Results in Sect. 3 either are independent of testbed location (3.1 Protocols), or have been validated in the PlanetLab (3.2 Data Centers). All results in Sect. 4 (Capabilities) and 5.3 (Overhead) are independent of testbed location as well. When discussing the remaining results (Sects. 5.1 and 5.2), we have carefully focused on "bottlenecks caused by the lack of client capabilities" and "the effects of data center placement". While these results might change for each specific service if measured from other locations, the core message in our conclusions (effects of design choices) still holds. We have changed several statements in the paper, following reviewers' suggestions, in order to make our claims clearer and give the right impression about the research.

The reviewers suggested that we could use a distributed testbed (e.g., the PlanetLab) to perform experiments. We indeed use the PlanetLab for our measurements in Sect. 3. We have discarded this option for the experiments in Sect. 5, because those experiments require a well-controlled environment, and the PlanetLab is unreliable for such performance measurements.

The reviewers requested more information about our measurement setup and chosen benchmarks. We have addressed these points by extending Sect. 2, which now gives more details about both our testbed and benchmark sets. Sect. 5 has been extended as well, providing more information about our results. In the same line, the paper now provides a URL from where readers can download both our benchmark scripts and all data evaluated in the paper: http://www.simpleweb.org/wiki/cloud_benchmarks

Reviewers commented that our discussions were predictable and resembled a product comparison. We believe this impression was created by the way Sect. 6 was organized in the previous manuscript. We have revised Sect. 6 to emphasize design choices and their respective consequences to performance, thus making our contributions clearer.

Finally, the reviewers recommended that we should take a neutral voice while describing results. We have followed the recommendation and changed the language/text accordingly in several parts of the text. Other minor corrections (e.g. on addressing previous works) also have been incorporated accordingly.