EdgeDroid 1.0

An Experimental Approach to Benchmarking Human-in-the-Loop Applications

ROYAL INSTITUTE OF TECHNOLOGY

Manuel Olguín, Junjue Wang,
Mahadev Satyanarayanan, James Gross

Carnegie Mellon University Computer Science Department

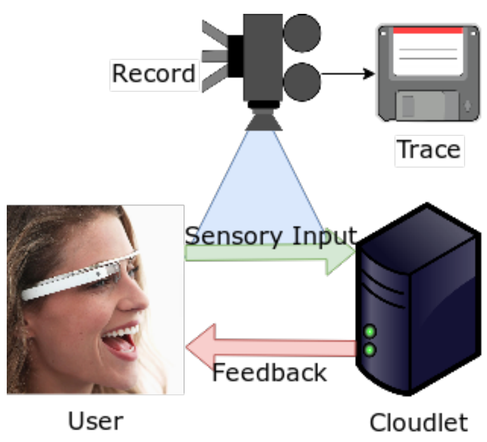
Abstract

Many emerging mobile applications, including augmented reality (AR) and wearable cognitive assistance, aim to provide seamless user interaction. However, the complexity of benchmarking these human-in-the-loop applications limits reproducibility and makes performance evaluation difficult. In this paper, we present EdgeDroid, a benchmarking suite designed to reproducibly evaluate these applications.

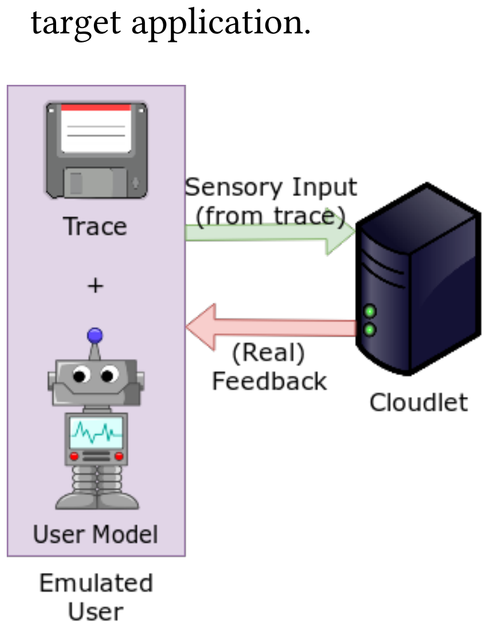
Our core idea rests on recording traces of user interaction, which are then replayed at benchmarking time in a controlled fashion based on an underlying model of human behavior. This allows for an automated system that greatly simplifies benchmarking large scale scenarios and stress testing the application. Our results show the benefits of EdgeDroid as a tool for both system designers and application developers.

Basic Idea

- Benchmarking human-in-the-loop applications is **hard** due to human users:
  - They are unpredictable.
  - They make scaling difficult (you need more of them!).
- What if we could cut out the user?



Step 1: Trace user input while operating the target application.



Step 2: Replace user with recorded trace plus a user model.

Design & Implementation

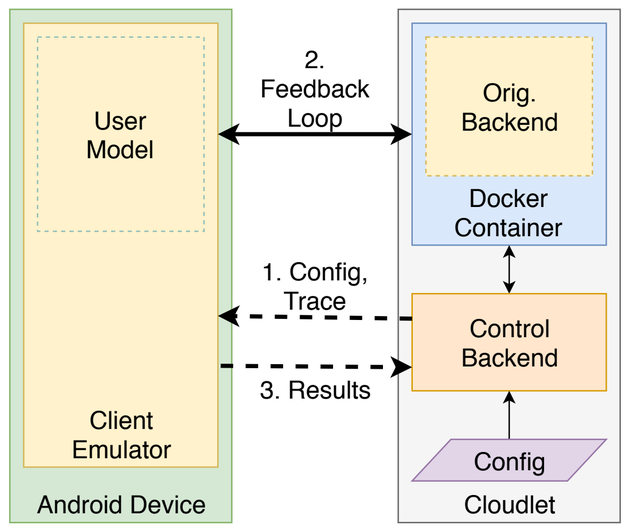


Figure 1: Suite Architecture

- The *control backend* controls the experiments and collects measurements from the application and the cloudlet itself. Implemented in Python 3.6.
- The *client emulators* play out a prerecorded sensory input trace over the network in a controlled fashion, while collecting relevant metrics. Implemented in Java using the Android SDK.

User Model

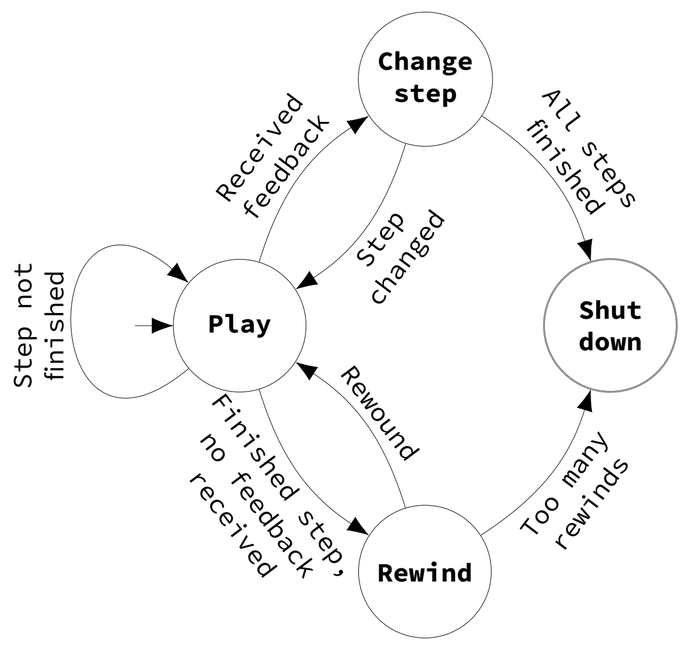
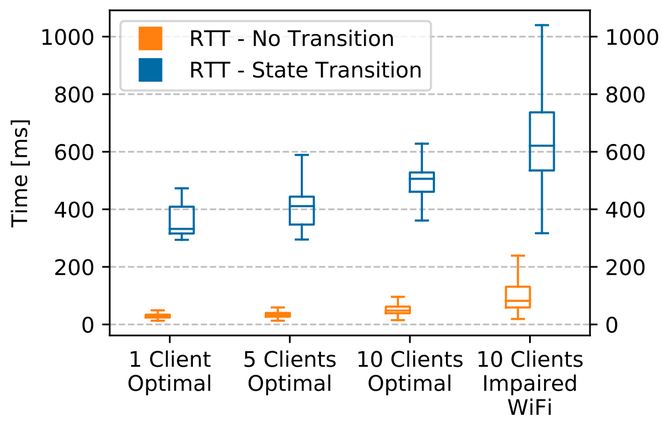


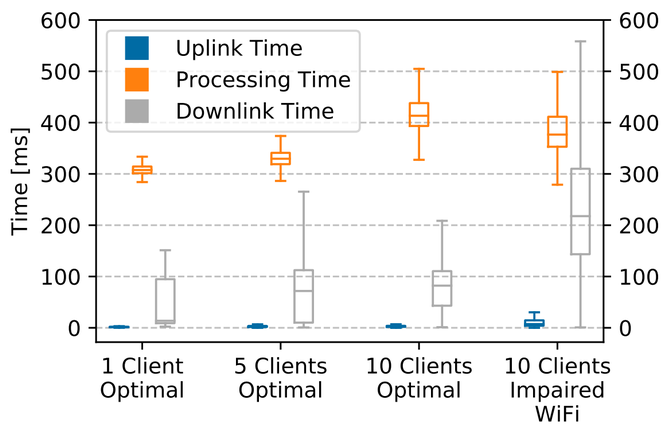
Figure 2: Model of human interaction.

We implement a very simple user model in order to be able to react to feedback from the application and adapt our replay of the trace to the current system conditions. In EdgeDroid 1.0, our model is that of a user who does not suffer any of the shortcomings of real human users such as annoyance, fatigue, frustration, nausea. In the future, we envision creating many versions of EdgeDroid (i.e., EdgeDroid 2.0, EdgeDroid 3.0, etc.) that embody more human-like user models that more accurately emulate attributes such as those mentioned above.

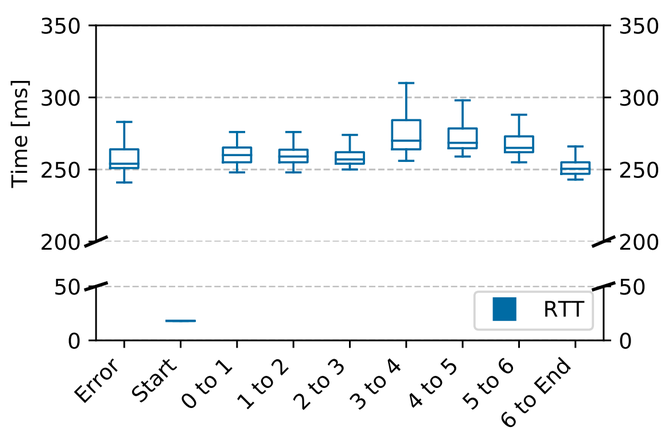
Some Example Results



a: Comparison of total RTTs.



b: RTTs by major pipeline segments.



c: RTTs by step transition.

Figure 3: Example results for a benchmark of the LEGO application in [1].

These results could be useful for:

- System designers wishing to identify bottlenecks in their hardware stack.
- Application developers who want to find potential for optimization.

References

[1] Z. Chen *et al.*, "Early Implementation Experience with Wearable Cognitive Assistance Applications," in *Proceedings of the 2015 Workshop on Wearable Systems and Applications*.

[2] K. Ha *et al.*, "Towards Wearable Cognitive Assistance," in *Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services*.

[3] Z. Chen *et al.*, "An Empirical Study of Latency in an Emerging Class of Edge Computing Applications for Wearable Cognitive Assistance," in *Proceedings of the Second ACM/IEEE Symposium on Edge Computing*.

[4] M. Satyanarayanan *et al.*, "The Case for VM-based Cloudlets in Mobile Computing," *IEEE Pervasive Computing*, vol. 8, no. 4, 2009.

[5] T. Bittmann, "The Edge Will Eat the Cloud," *Gartner Research*, 2017.