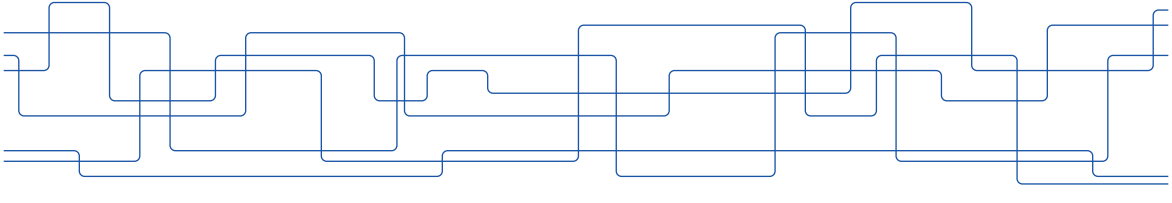
EdgeDroid

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

M. Olguín Muñoz[†], J. Wang[‡], M. Satyanarayanan[‡] and J. Gross[†]

[†] KTH Royal Institute of Technology
[‡] Carnegie Mellon University

HotMobile'19 Session 5: February 28th 2019, Santa Cruz, CA




65
126ft
NE
354


SAMSUNG

Sensory Input

Human-parseable
Feedback

Studying Human-in-the-Loop Applications

Need to understand and optimize these applications:

- How do they interact with each other?
- How do they interact with infrastructure?
- How do they scale?

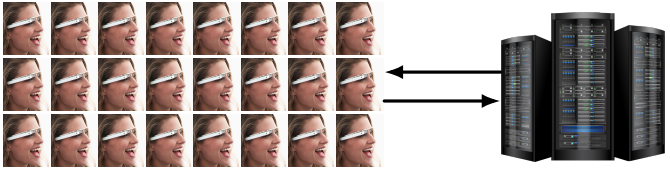
With which methodology can we study these behaviors?

Studying Human-in-the-Loop Applications

Need to understand and optimize these applications:

- How do they interact with each other?
- How do they interact with infrastructure?
- How do they scale?

With which methodology can we study these behaviors?

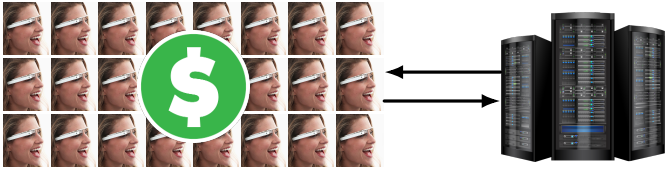


Studying Human-in-the-Loop Applications

Need to understand and optimize these applications:

- How do they interact with each other?
- How do they interact with infrastructure?
- How do they scale?

With which methodology can we study these behaviors?



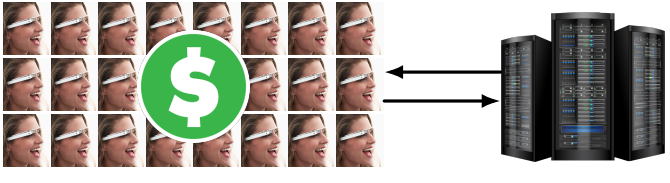
Costly

Studying Human-in-the-Loop Applications

Need to understand and optimize these applications:

- How do they interact with each other?
- How do they interact with infrastructure?
- How do they scale?

With which methodology can we study these behaviors?



Costly, poor repeatability

Previous & Related Work

Prototypes

Latencies

Modeling

Our Contributions

Previous & Related Work

Prototypes Latencies Modeling

Our Contributions

- A methodology for benchmarking human-in-the-loop applications.

Previous & Related Work

Prototypes Latencies Modeling

Our Contributions

- A methodology for benchmarking human-in-the-loop applications.
- EdgeDroid: A benchmarking tool-suite.

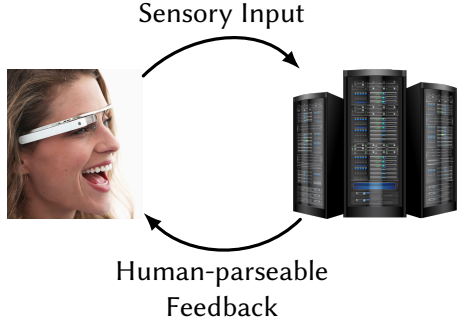
Previous & Related Work

Prototypes Latencies Modeling

Our Contributions

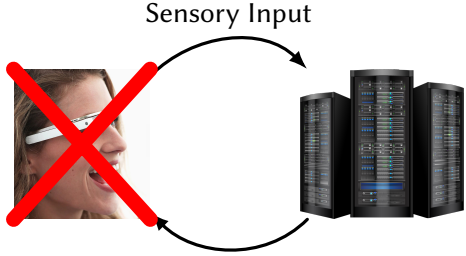
- A methodology for benchmarking human-in-the-loop applications.
- EdgeDroid: A benchmarking tool-suite.
- Experiments and measurements which show the effectiveness of the approach.

Approach: Motivation



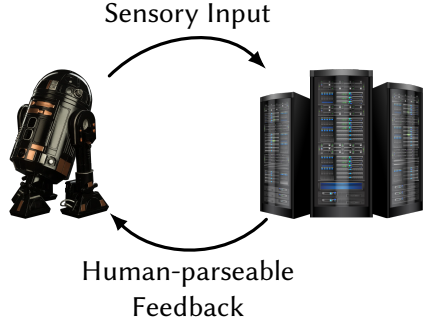
Benchmarking human-in-the-loop applications is HARD

Approach: Motivation



What if we could do away with the human users?

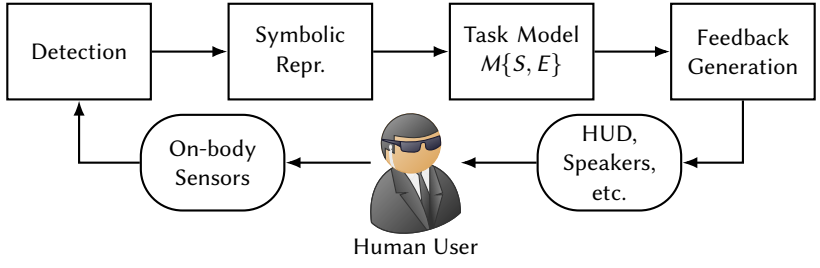
Approach: Motivation



What if we could do away with the human users?

Task-guidance Cognitive Assistance

Some image of task guidance: “put the blue lego on top of the red one”



Requirements

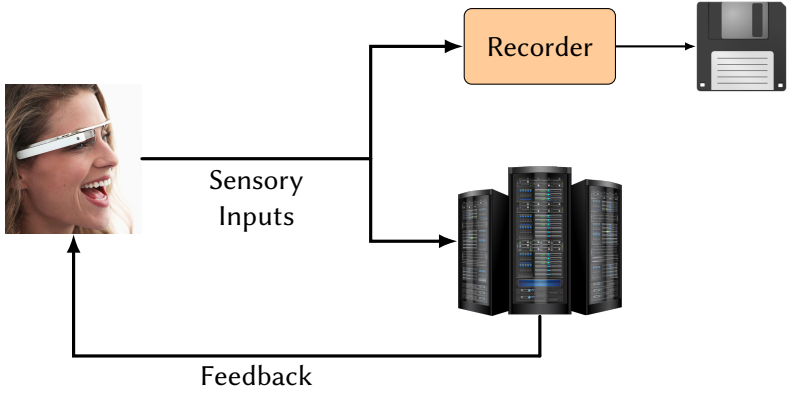
- Generate realistic, high-dimensional, real-time inputs.
- Correctly and realistically react to feedback.
- KPI: Delays.

Requirements

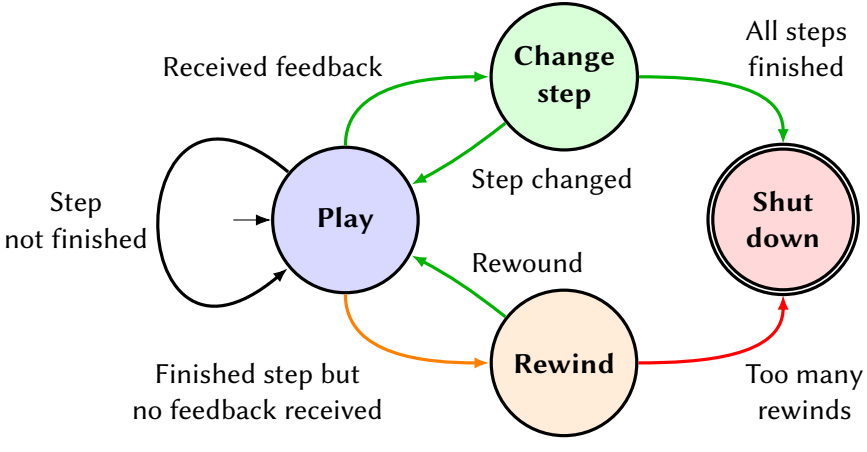
- Generate realistic, high-dimensional, real-time inputs.
- Correctly and realistically react to feedback.
- KPI: Delays.

Trace of pre-recorded inputs & a model of user behavior

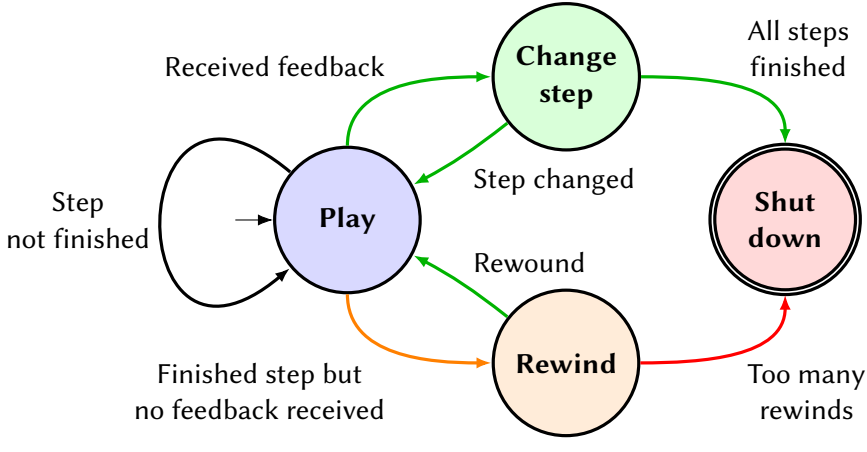
Tracing



User Model

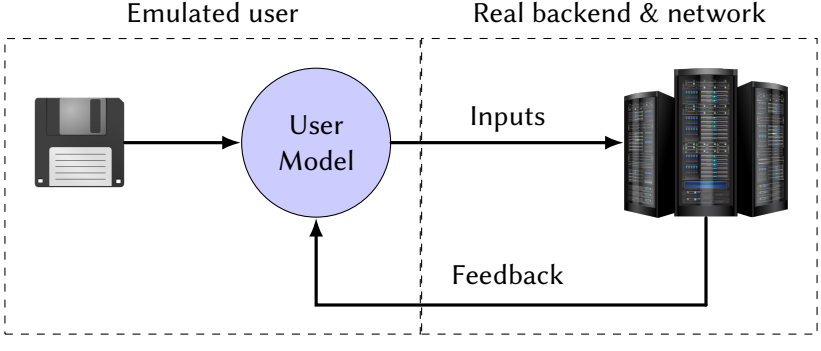

Change step
All steps finished
Received feedback
Step changed
Step not finished
Play
Shut down
Rewound
Rewind
Finished step but no feedback received
Too many rewinds

User Model

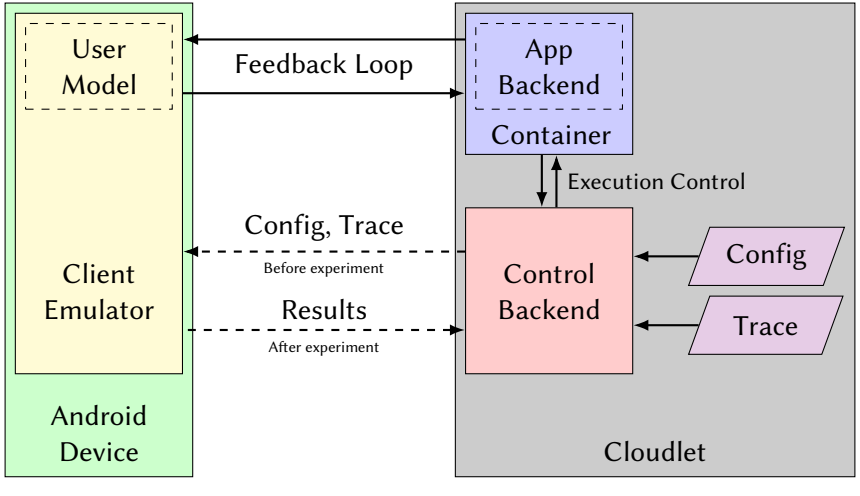


Currently working on a more thorough characterization of human behavior.

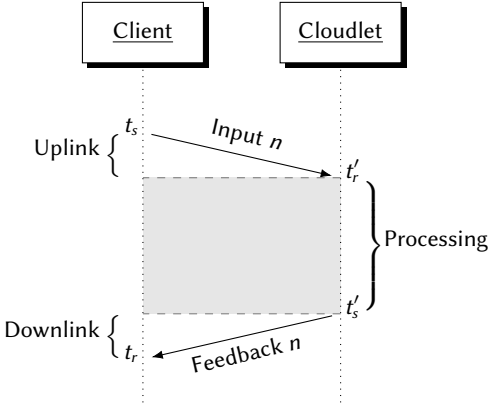
Trace Replay



Implementation



Timestamping



Clocks are synchronized previous to the experiment.

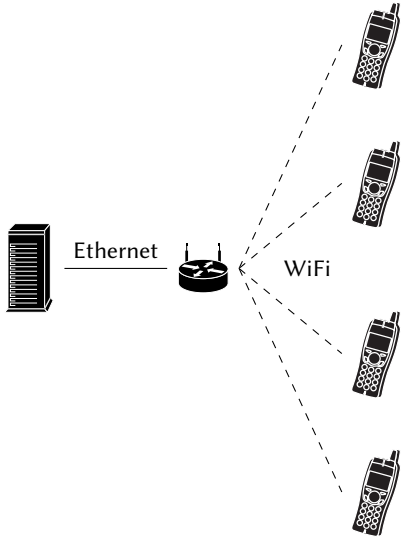
Timestamps at key points to obtain:

$$\Delta T_{\text{up}} = t'_r - t_s \tag{1}$$
$$\Delta T_{\text{proc}} = t'_s - t'_r \tag{2}$$
$$\Delta T_{\text{down}} = t_r - t'_s \tag{3}$$

$$\Delta T_{\text{total}} = \Delta T_{\text{up}} + \Delta T_{\text{proc}} + \Delta T_{\text{down}} = t_r - t_s \tag{4}$$

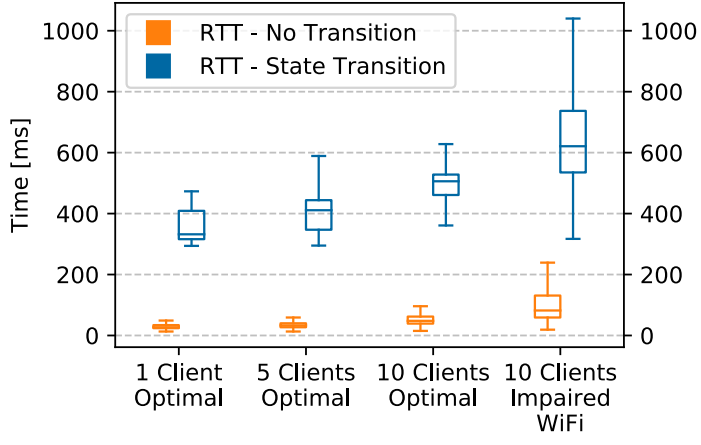
Evaluation

Insert pictures of LEGO Assistant



Results

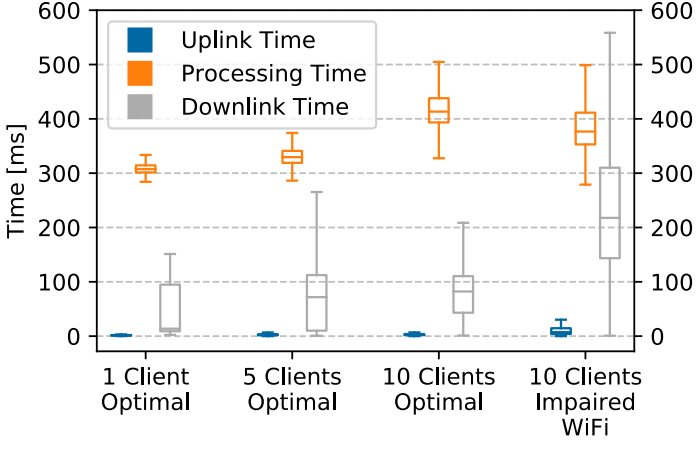
Figure labels



I haven't explained the task model, maybe skip this graph?

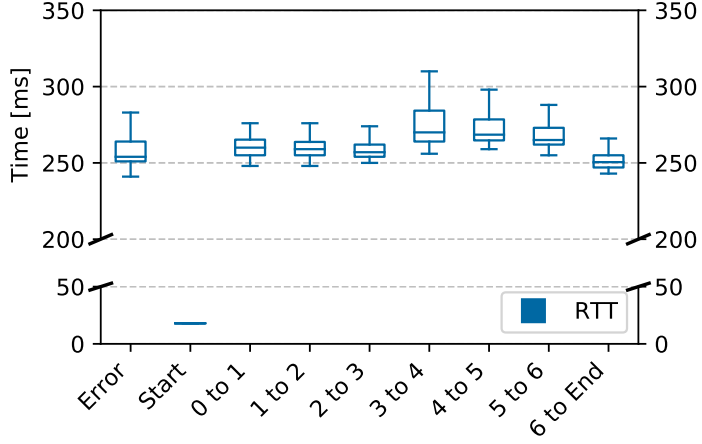
Results

Figure labels



Results

Figure labels



Future Work

- More accurate user model.
- Expand to other types of Applications.

Summary

- There's a need to study the scaling of Human-in-the-Loop applications.
  - This is difficult due to human users.
- We present a methodology + tool suite for benchmarking:
  - **EdgeDroid**
  - Trace based.
  - Model of human behavior.
- We present results which show the utility of EdgeDroid.