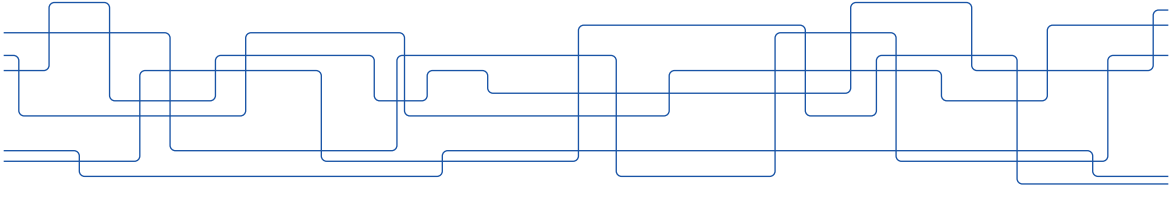
EdgeDroid

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?

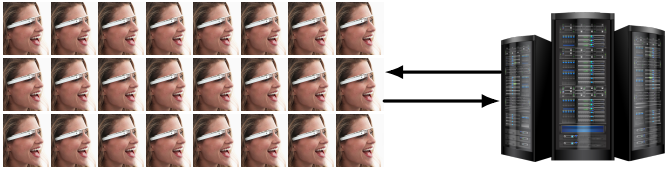
With which methodology can we study these behaviors?

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?

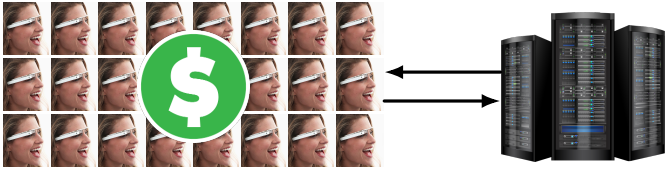


Studying Human-in-the-Loop Applications

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?



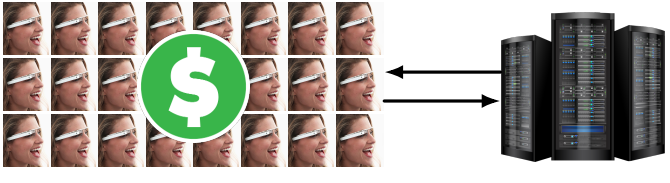
Costly

Studying Human-in-the-Loop Applications

Need to understand and optimize these applications:

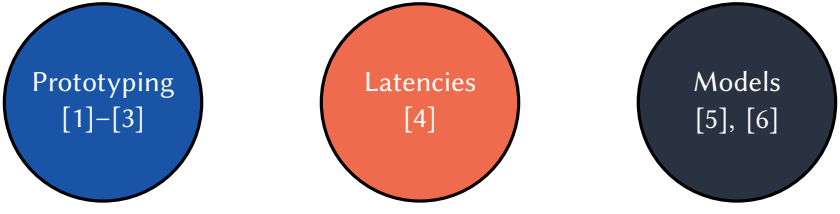
- How do they interact with each other?
- How do they interact with infrastructure?
- How do they scale?

With which methodology can we study these behaviors?

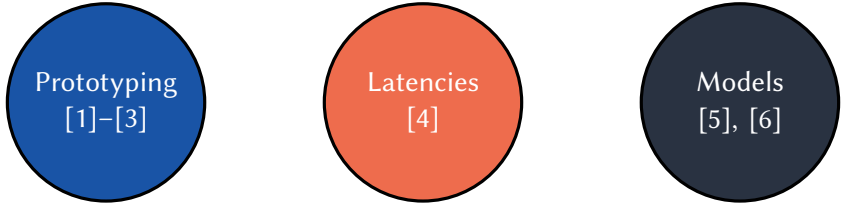


Costly, poor repeatability

Previous & Related Work



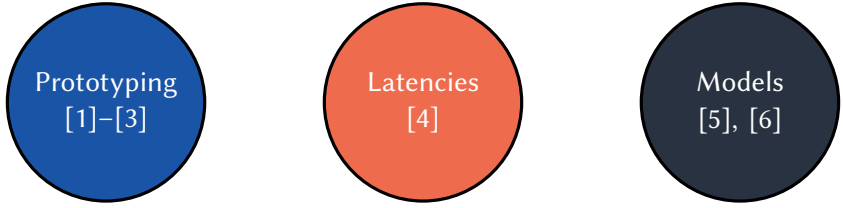
Previous & Related Work



Our Contributions

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

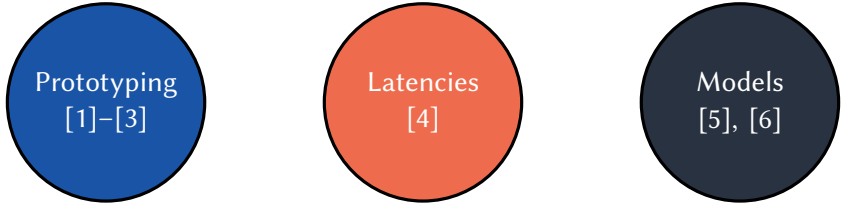
Previous & Related Work



Our Contributions

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

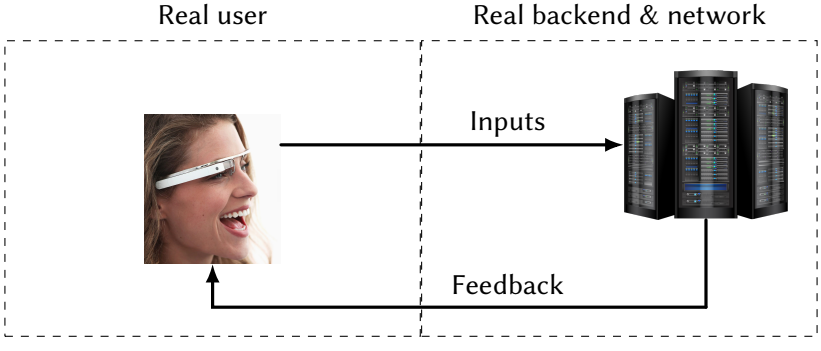
Previous & Related Work



Our Contributions

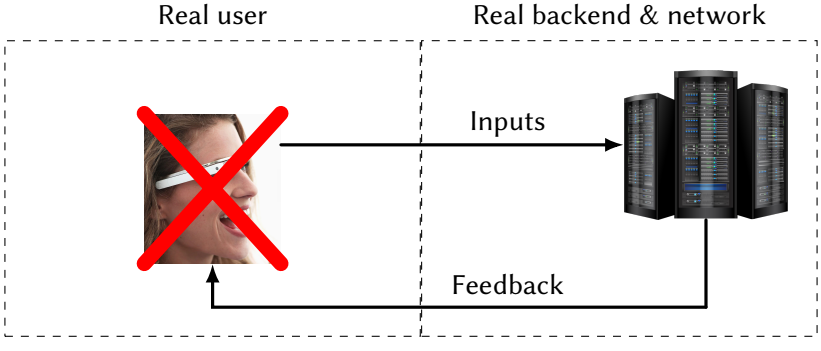
- A methodology for benchmarking human-in-the-loop applications.
- EdgeDroid: A benchmarking tool-suite.
- Experiments and measurements which show the effectiveness of the approach.

Approach



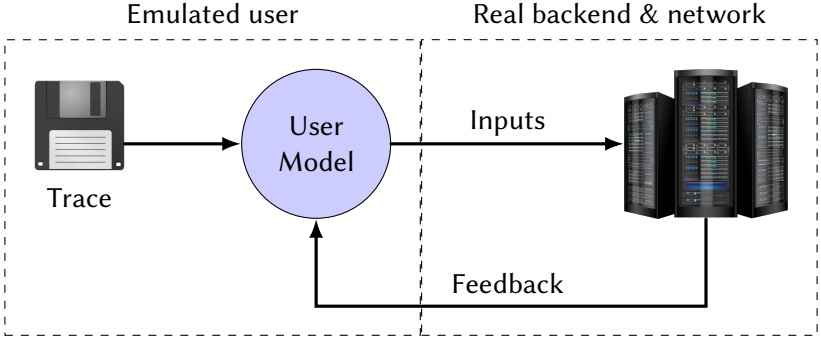
Benchmarking human-in-the-loop applications is HARD

Approach



What if we could do away with the human users?

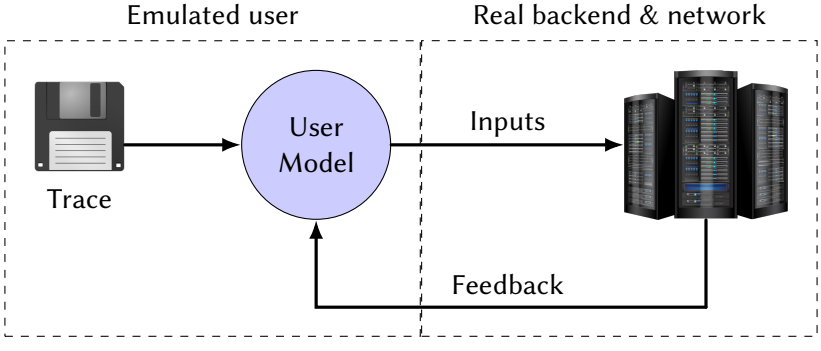
Approach



What if we could do away with the human users?

Repeatable, scalable!

Approach

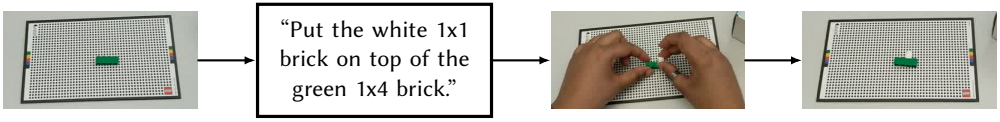


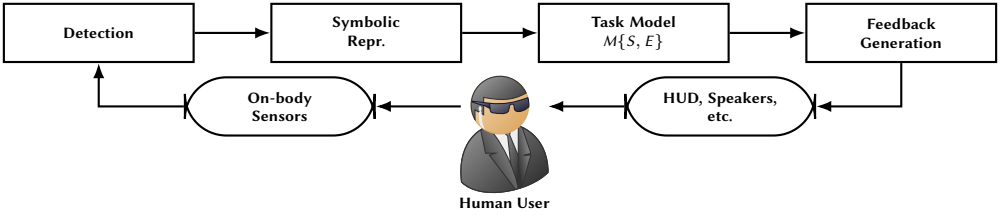
What if we could do away with the human users?

Repeatable, scalable!

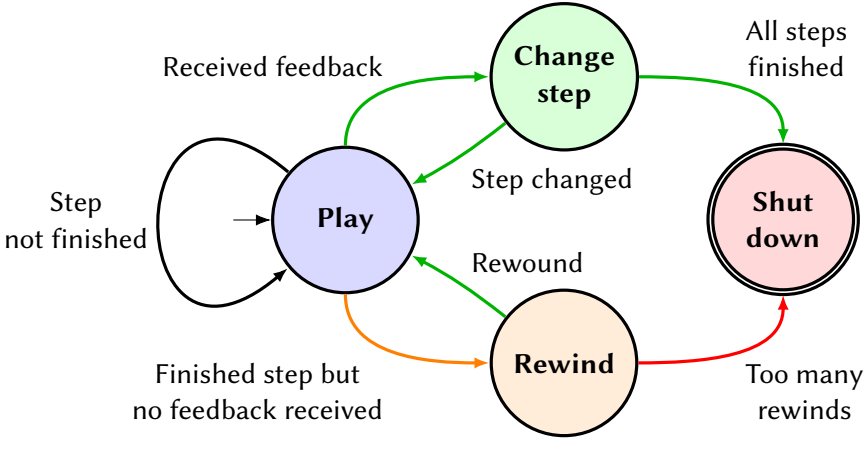
Key question: Credibility.

Task-guidance Wearable Cognitive Assistance [1]

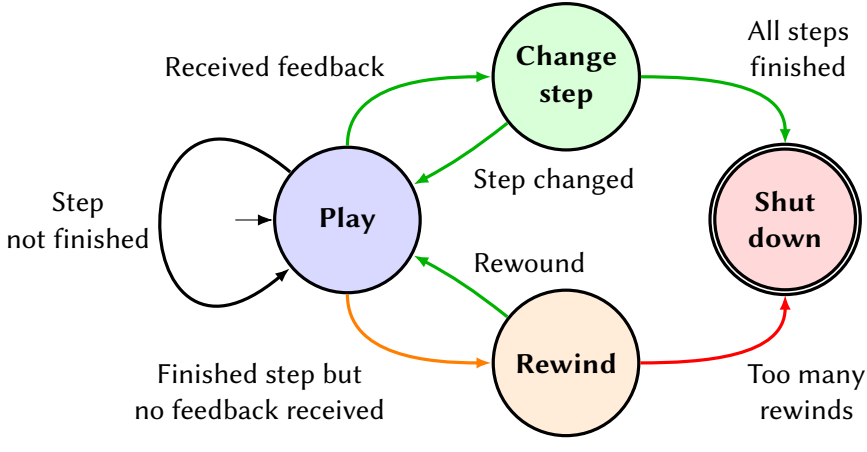




User Model

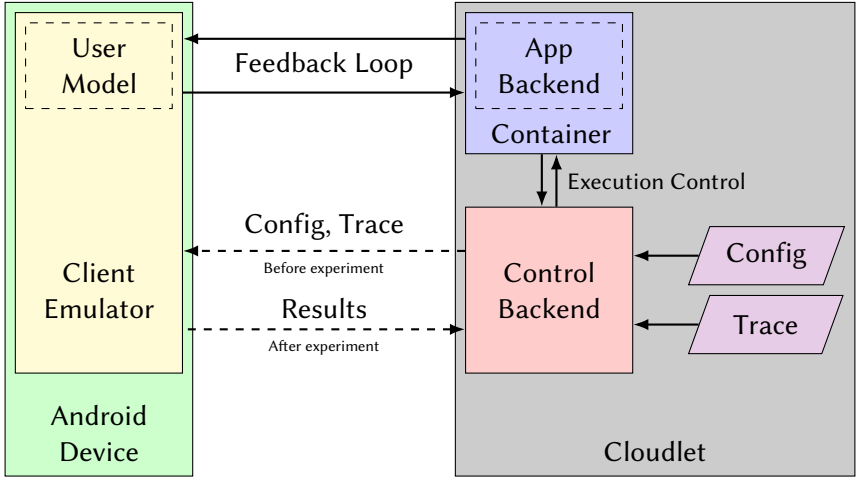

Change step
All steps finished
Received feedback
Step changed
Step not finished
Play
Shut down
Rewound
Rewind
Finished step but no feedback received
Too many rewinds

User Model



Currently working on a more thorough characterization of human behavior.

Implementation



https://github.com/molguin92/EdgeDroid

Evaluation

Key purpose:
Demonstrate utility of EdgeDroid.

Scenarios

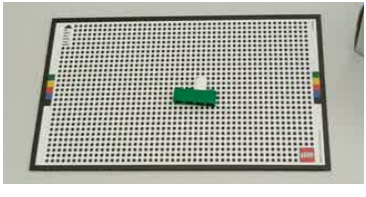
- *Optimal* scenarios with 1, 5 and 10 devices.
- Weakened wireless link with 10 devices.

Table: Latency bounds (Chen *et al.* [4]).

Latency [ms]	Quality
< 600	Excellent
$600 - 2700$	Impaired
> 2700	Unusable

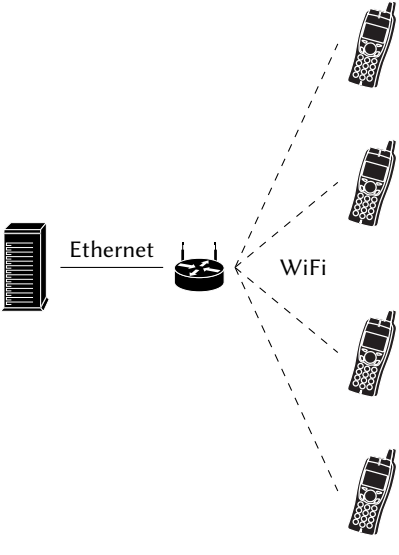
Evaluation Setup

Application



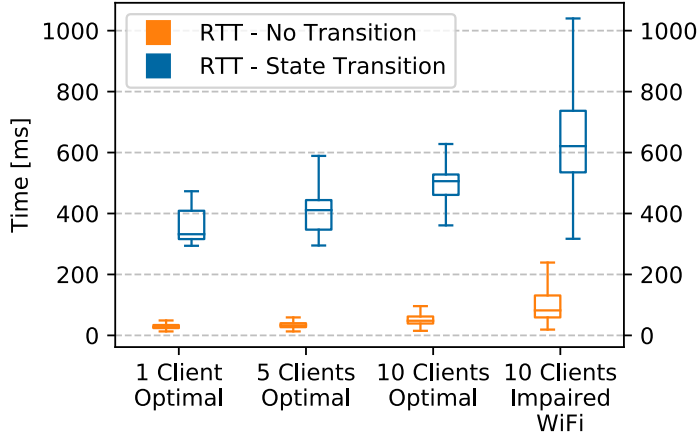
LEGO Assistant

- 7 steps
- 2 min average duration

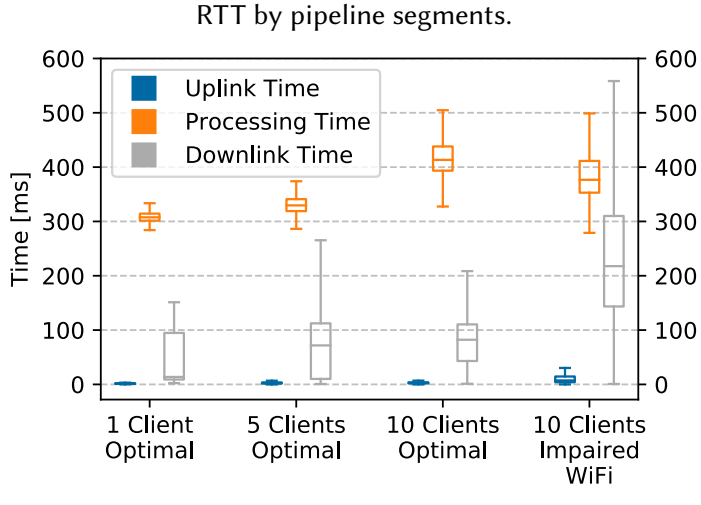


Results

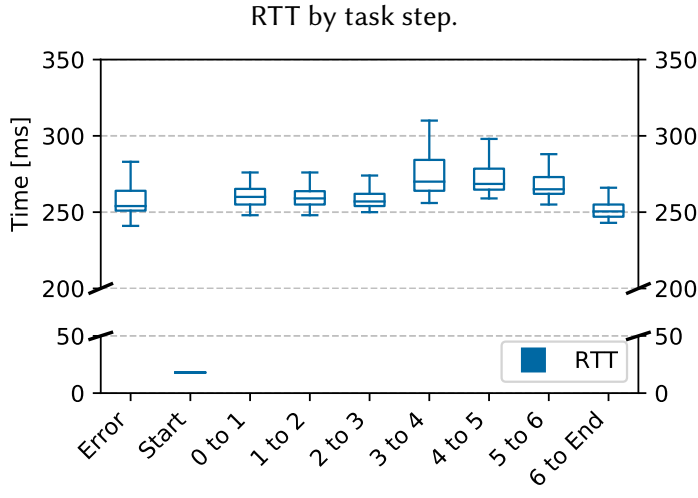
Inputs that triggered a state change vs. inputs that did not.



Results


RTT by pipeline segments.
Uplink Time
Processing Time
Downlink Time
Time [ms]
600
500
400
300
200
100
0
1 Client Optimal
5 Clients Optimal
10 Clients Optimal
10 Clients Impaired WiFi

Results


RTT by task step.
Time [ms]
350
300
250
200
50
0
RTT
Error
Start
0 to 1
1 to 2
2 to 3
3 to 4
4 to 5
5 to 6
6 to End

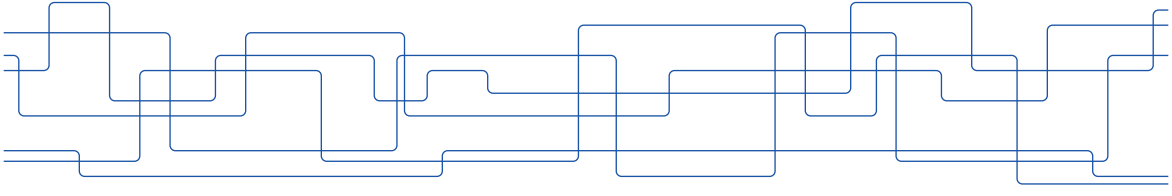
Conclusions

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.

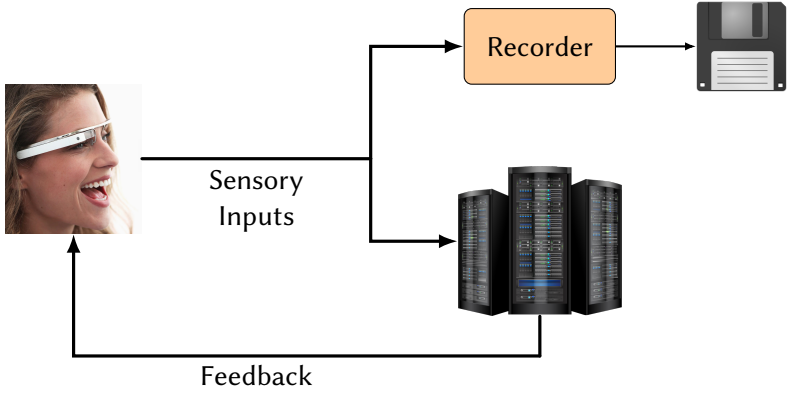


Requirements

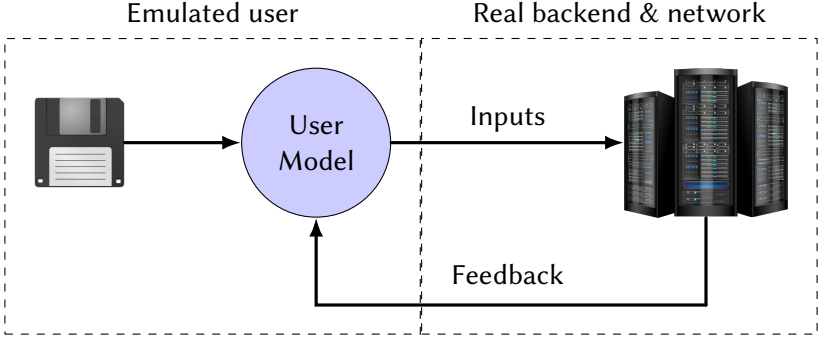
- Generate realistic, high-dimensional, real-time inputs.
- Correctly and realistically react to feedback.
- KPI: Delays.

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

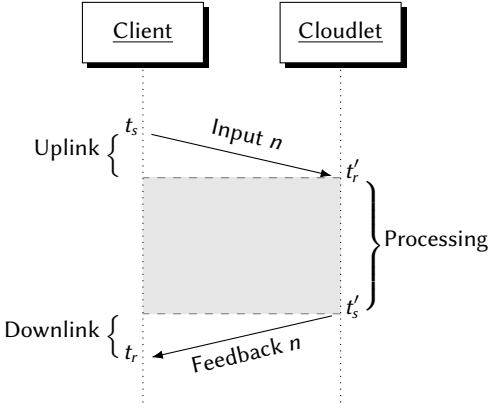
Tracing



Trace Replay



Timestamping



Clocks are synchronized previous to the experiment.

Timestamps at key points to obtain:

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

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

References I

K. Ha *et al.*, "Towards wearable cognitive assistance," in *Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '14, Bretton Woods, New Hampshire, USA: ACM, 2014, pp. 68–81, ISBN: 978-1-4503-2793-0. DOI: 10.1145/2594368.2594383. [Online]. Available: http://doi.acm.org/10.1145/2594368.2594383.

Z. Chen *et al.*, "Early implementation experience with wearable cognitive assistance applications," in *Proceedings of the 2015 Workshop on Wearable Systems and Applications*, ser. WearSys '15, Florence, Italy: ACM, 2015, pp. 33–38, ISBN: 978-1-4503-3500-3. DOI: 10.1145/2753509.2753517. [Online]. Available: http://doi.acm.org/10.1145/2753509.2753517.

D. Chatzopoulos *et al.*, "Hyperion: A wearable augmented reality system for text extraction and manipulation in the air," in *Proceedings of the 8th ACM on Multimedia Systems Conference*, ser. MMSys'17, Taipei, Taiwan: ACM, 2017, pp. 284–295, ISBN: 978-1-4503-5002-0. DOI: 10.1145/3083187.3084017. [Online]. Available: http://doi.acm.org/10.1145/3083187.3084017.

References II

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*, ser. SEC '17, San Jose, California: ACM, 2017, 14:1–14:14, ISBN: 978-1-4503-5087-7. DOI: 10.1145/3132211.3134458. [Online]. Available: http://doi.acm.org/10.1145/3132211.3134458.

H. Al-Zubaidy *et al.*, "Performance of in-network processing for visual analysis in wireless sensor networks," in *Proceedings of the IFIP Networking Conference*, ser. IFIP NETWORKING'15, 2015.

S. Schiessl *et al.*, "Finite-length coding in edge computing scenarios," in *Proceedings of the International Workshop on Smart Antennas*, ser. ITG WSA '17, 2017.

M. Satyanarayanan *et al.*, "The case for VM-based cloudlets in mobile computing," *IEEE Pervasive Computing*, vol. 8, no. 4, 2009.

J. Flinn, "Cyber foraging: Bridging mobile and cloud computing," *Synthesis Lectures on Mobile and Pervasive Computing*, vol. 7, no. 2, pp. 1–103, 2012.

K. Sasaki *et al.*, "Vehicle control system coordinated between cloud and mobile edge computing," in *2016 55th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*, 2016, pp. 1122–1127. DOI: 10.1109/SICE.2016.7749210.

References III

——,"Layered vehicle control system coordinated between multiple edge servers," in *2017 IEEE Conference on Network Softwarization (NetSoft)*, 2017, pp. 1–5. DOI: 10.1109/NETSOFT.2017.8004199.

T. Bittmann, "The edge will eat the cloud," *Gartner Research*, no. G00338633, 2017.

K. Kumar *et al.*, "Cloud computing for mobile users: Can offloading computation save energy?" *IEEE Computer*, vol. 43, no. 4, pp. 51–56, 2010.

E. Cuervo *et al.*, "Maui: Making smartphones last longer with code offload," in *Proceedings of the International Conference on Mobile Systems, Applications, and Services*, ser. ACM MOBISYS'10, 2010.

K. Ha *et al.*, "The impact of mobile multimedia applications on data center consolidation," in *2013 IEEE International Conference on Cloud Engineering (IC2E)*, 2013, pp. 166–176. DOI: 10.1109/IC2E.2013.17.

K. Ha *et al.*, "Just-in-time provisioning for cyber foraging," in *Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '13, Taipei, Taiwan: ACM, 2013, pp. 153–166, ISBN: 978-1-4503-1672-9. DOI: 10.1145/2462456.2464451. [Online]. Available: http://doi.acm.org/10.1145/2462456.2464451.

(2018). Docker, [Online; accessed 14. Aug. 2018], [Online]. Available: https://www.docker.com.

References IV

(2018). Network Time Protocol, [Online; accessed 24. Sep. 2018], [Online]. Available: https://www.eecis.udel.edu/~mills/ntp/html/index.html.

(2018). TOML, [Online; accessed 25. Sep. 2018], [Online]. Available: https://github.com/toml-lang/tomll.

K. Kim *et al.*, "Workload synthesis: Generating benchmark workloads from statistical execution profile," in *2014 IEEE International Symposium on Workload Characterization (IISWC)*, 2014, pp. 120–129. DOI: 10.1109/IISWC.2014.6983051.

E. Deniz *et al.*, "Minime: Pattern-aware multicore benchmark synthesizer," *IEEE Transactions on Computers*, vol. 64, no. 8, pp. 2239–2252, 2015, ISSN: 0018-9340. DOI: 10.1109/TC.2014.2349522.

M. Olguín *et al.*, "Demo: Scaling on the Edge – A Benchmarking Suite for Human-in-the-Loop Applications," in *Proceedings of The Third ACM/IEEE Symposium on Edge Computing*, ser. SEC '18, Accepted Submission, Extended Abstract, 2018. [Online]. Available: https://olguin.se/files/demo-scaling-edge.pdf.