

Research Methods in computer science

Fall 2021

Lecture 24

Omprakash Gnawali

November 15, 2021

Agenda

Experience writing a paper

Sampling Bias

Conference Updates

Experiences Writing

Please submit a response covering these topics

- Your prior experience writing a paper
- What 1-2 ideas/techniques helped you improve your writing
- What 1-2 ideas/techniques did not help as much as you would have liked
- Overall how do you rate your writing skills

[10 mins]

Sampling Bias

“In statistics, sampling bias is a bias in which a sample is collected in such a way that some members of the intended population are less likely to be included than others. It results in a biased sample, a non-random sample of a population (or non-human factors) in which all individuals, or instances, were not equally likely to have been selected. If this is not accounted for, results can be erroneously attributed to the phenomenon under study rather than to the method of sampling.” -- wikipedia

Types of sampling bias

Self selection bias

Pre-screening

Exclusion

etc.

[from wikipedia]

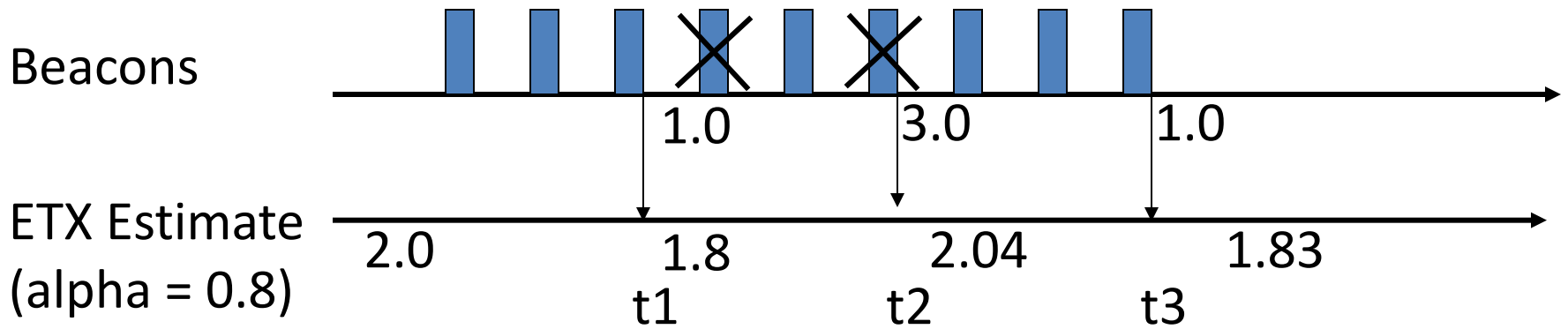
Many of you do research in AI/ML and may be aware of sampling bias in that context. It will be interesting to see examples of sampling bias in systems research.

Using signal strength for link
quality estimation can
introduce sampling bias.

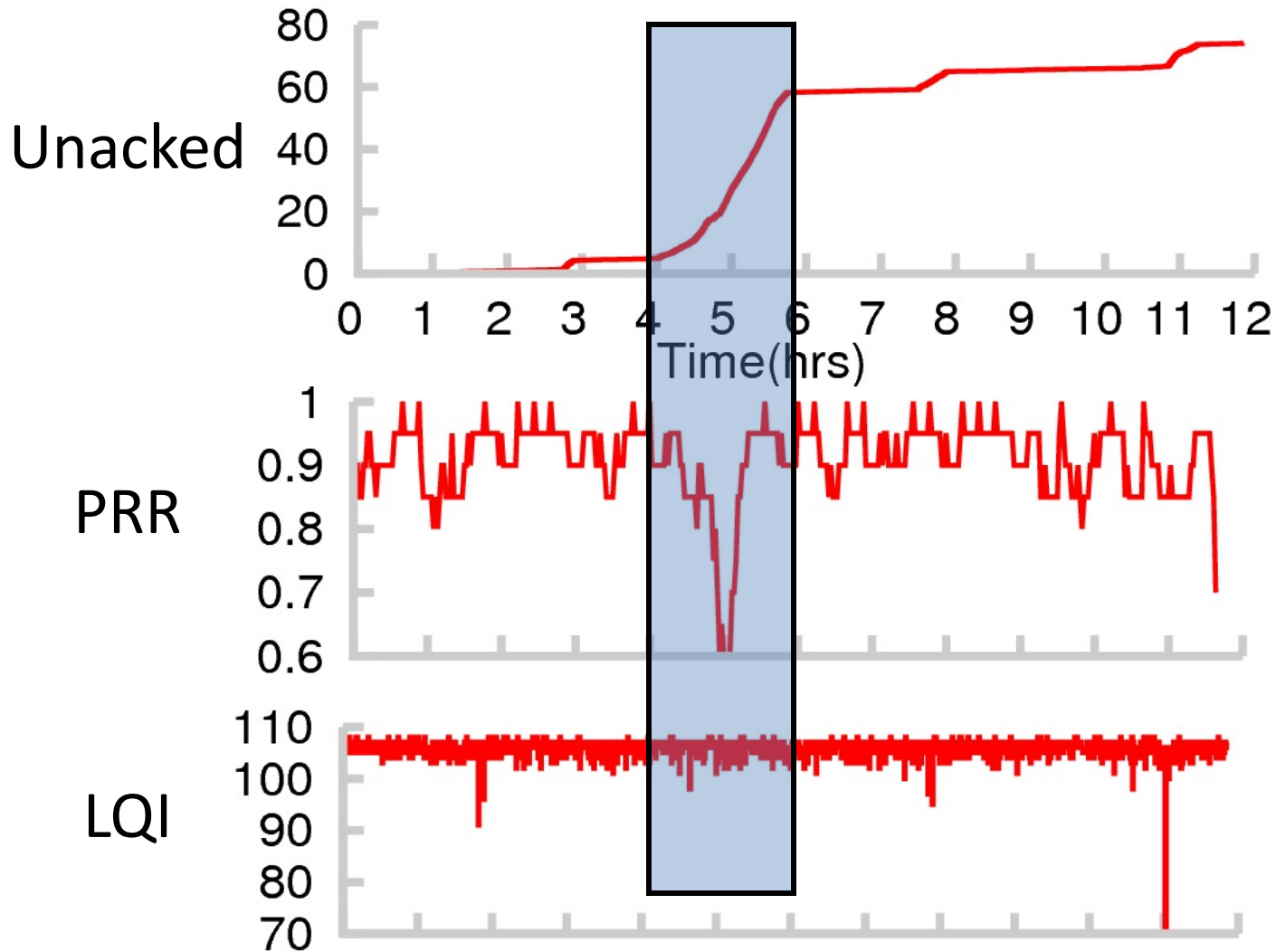
Link quality estimation

Estimate how “good” a link is.
Important for link selection.

ETX Estimation Example



Link Estimation using PHY info

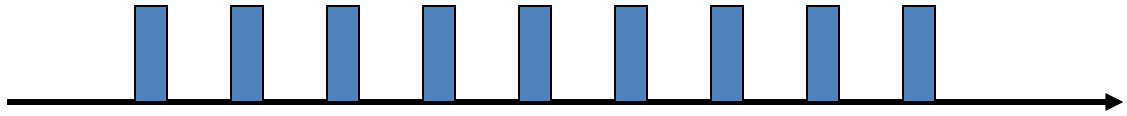


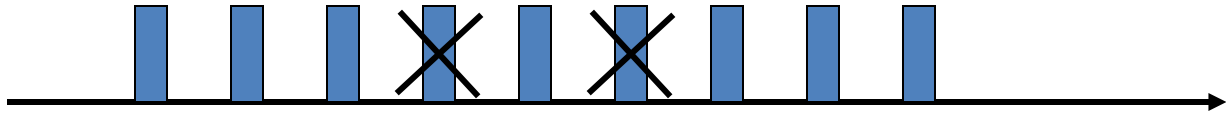
Quality of reception = Signal / Noise

Lot of wireless network research tries to understand performance as some function of SNR

“The Prism 2.5 chip-set provides per-frame measurements called RSSI (receive signal strength indication) and “silence value.” The RSSI reflects the total power observed by the radio hardware while receiving the frame, including signal, interference, and background noise. The silence value reflects the total power observed just before the start of the frame. We found that the accuracy of the RSSI and silence readings was within 4 dB by comparison with a spectrum analyzer. This paper reports signal-to-noise ratios derived from the RSSI and silence values.” – [Aguayo et al. 2004]

Quality of reception = Signal / Noise



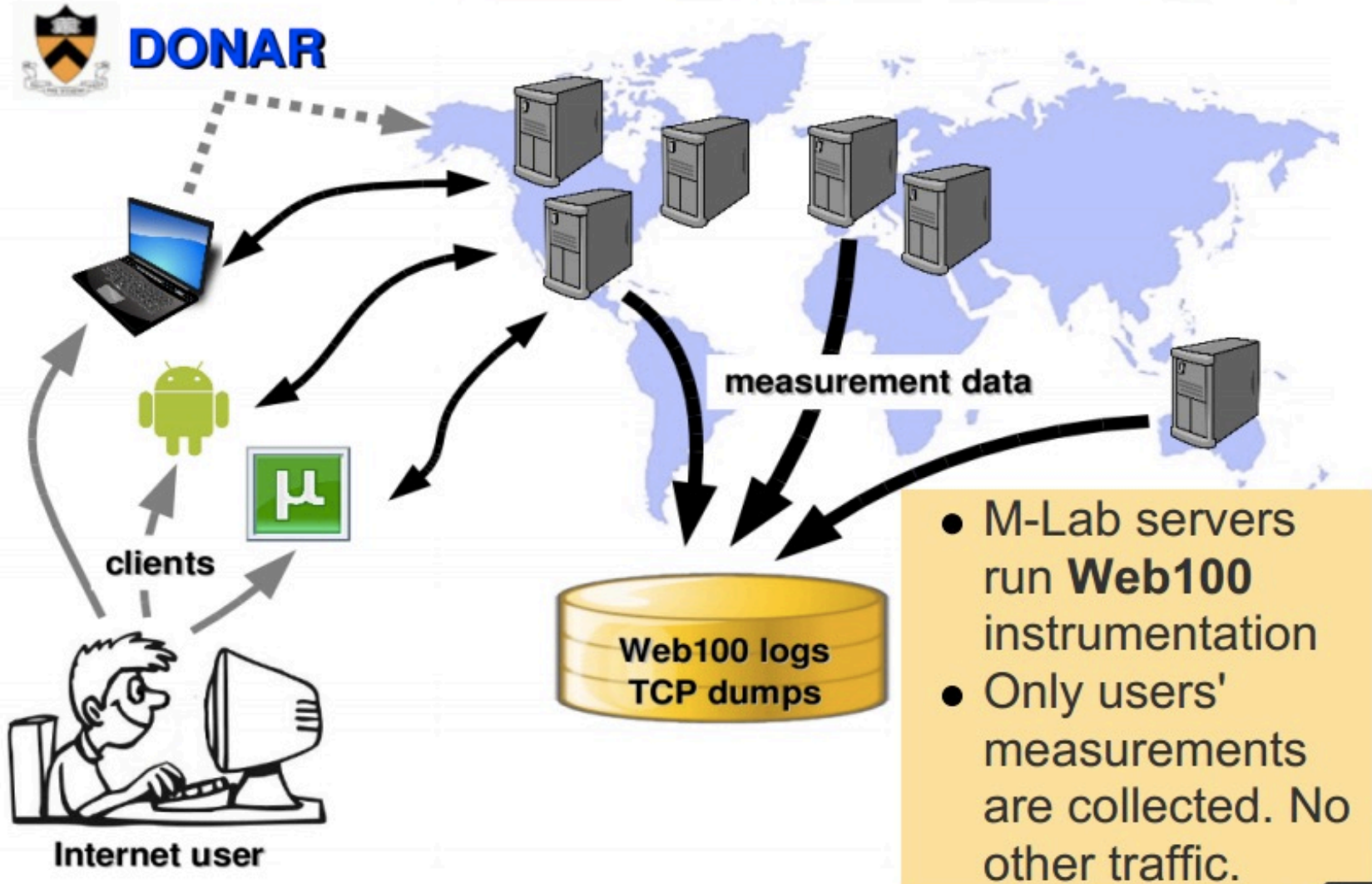


How to select participants
for an HCI study?

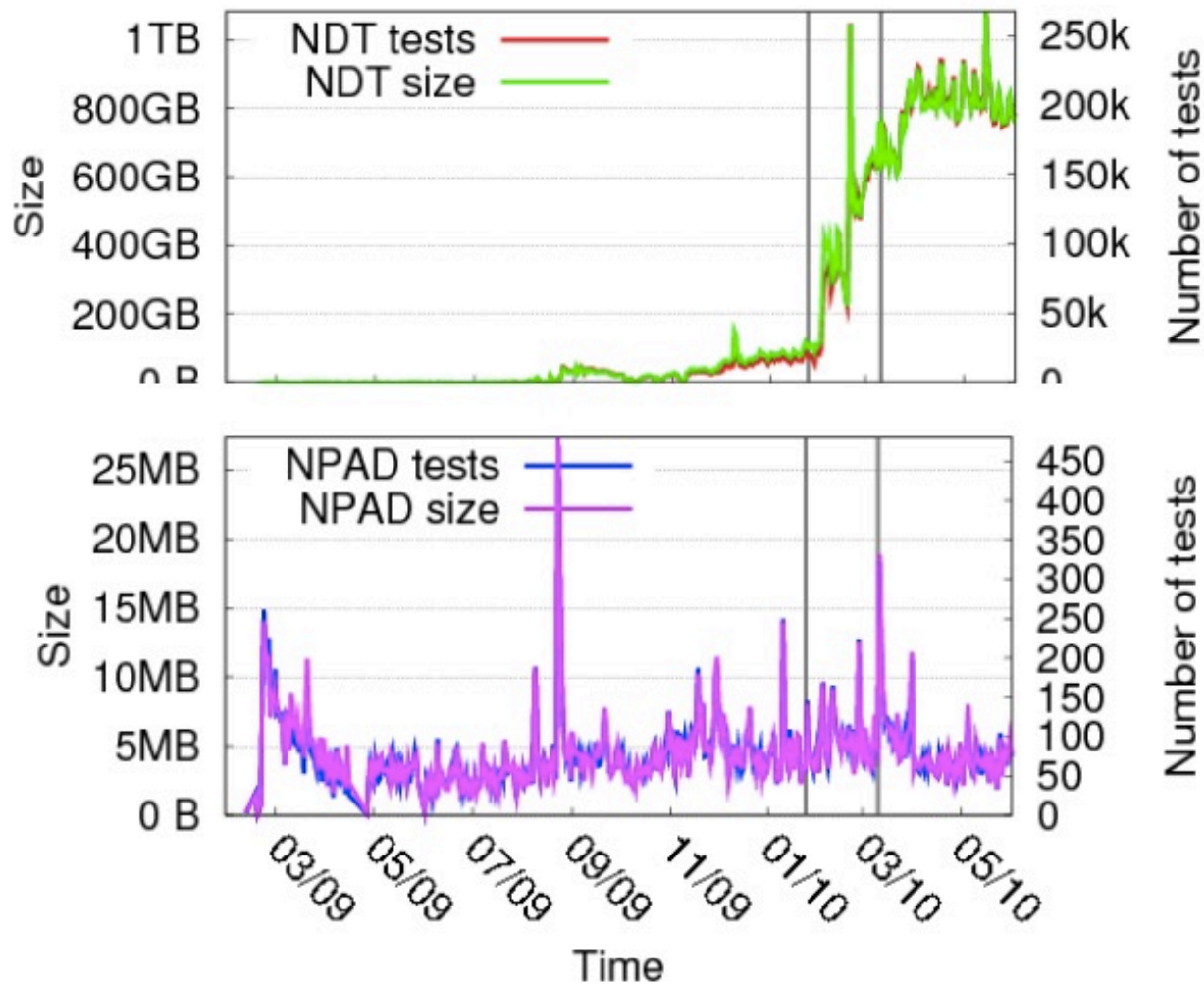
What is the possibility of
sampling bias?

Example of bias due to a
significant change in mixture of
data source

Measurement & Data collection



How much data? How many tests?



Jan 25 2010
uTorrent launch
Mar 11 2010
FCC launch

NDT

Tot tests: 22M

Tot size: 93TB

NPAD

Tot tests: 34K

Tot size: 2GB

Dataset Bias in Object Recognition Research

Unbiased Look at Dataset Bias [CVPR 2011]

Object Recognition Research

Dataset is a set of pictures of objects

Run algorithm to recognize/identify objects

Compute accuracy or other metrics

What are potential dataset bias?

Sampling, Capture, Negative Set

How to reduce selection bias
in visual object recognition datasets?

Research that uses online/social media data

Research: how people communicate, spread information, discuss, decide, etc.

What are some potential bias in the dataset?

What are the implications?

Exercise

Think of a ML/big-data idea

Identify a dataset on which you want to do machine learning

Describe a potential bias there in the application

Paper Review

You are expected to be working on paper review

Confirm that you are able to see papers assigned to you

Typical Template

Summary

Strengths

Weaknesses

Detailed Comments

Justification for these sections?

Review considerations

- Technical Claims
 - What are the claims?
 - Have the claims been validated?
 - Experimental
 - Data
 - Theory
- Technical Correctness

Conference Updates

Paper review submissions coming up

Other logistics

HW 11

Review the papers you were assigned

Summary

Strengths

Weaknesses

Details