Research Methods in computer science

Spring 2019

Lecture 17

Omprakash Gnawali

March 20, 2019

Agenda

HW7 live grading

Reading Papers and Paper expecations

Sampling Bias

HW8

Paper Expectations

Readers and reviewers set expectations

What are they?

Fair and unfair expectations.

Expectations are related to
how people say they read papers

How to Read a Scientific Paper

Begin with introduction, not abstract.
Identify the big question
Summarize the background in five sentences
Identify the specific questions
Identify the approach
Read the methods section
Read the results section
Determine if the results answer the questions
Read the conclusions/discussion/interpretation section
Read the abstract
Find out what others say about the paper

https://www.huffingtonpost.com/jennifer-raff/how-to-read-and-understand-a-scientific-paper_b_5501628.html

How to read a research paper

Goal is to understand the scientific contribution

Read critically

- Question the study, approach, …

Read creatively

- Extrapolate, extend, generalize, …

Make notes

Summarize

Compare

https://www.eecs.harvard.edu/~michaelm/postscripts/ReadPaper.pdf

How to Read a Paper

First pass [5-10 mins]

High level idea, category, context, contributions

Second pass [1 hr]

Some results, key ideas of the paper and key evidence

Third pass [variable]

Attention to deal, re-create the paper

http://ccr.sigcomm.org/online/files/p83-keshavA.pdf

How to Read an Engineering Research Paper

Read to answer questions

1. What are motivations for this work?
2. What is the proposed solution?
3. What is the work's evaluation of the proposed solution?
4. What is your analysis of the identified problem, idea and evaluation?
5. What are the contributions?
6. What are future directions for this research?
7. What questions are you left with?
8. What is your take-away message from this paper?

Paper Notes

Things worth remembering

Results, Ideas, Authors, ….

Electronic systems

Could be integrated with References

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.

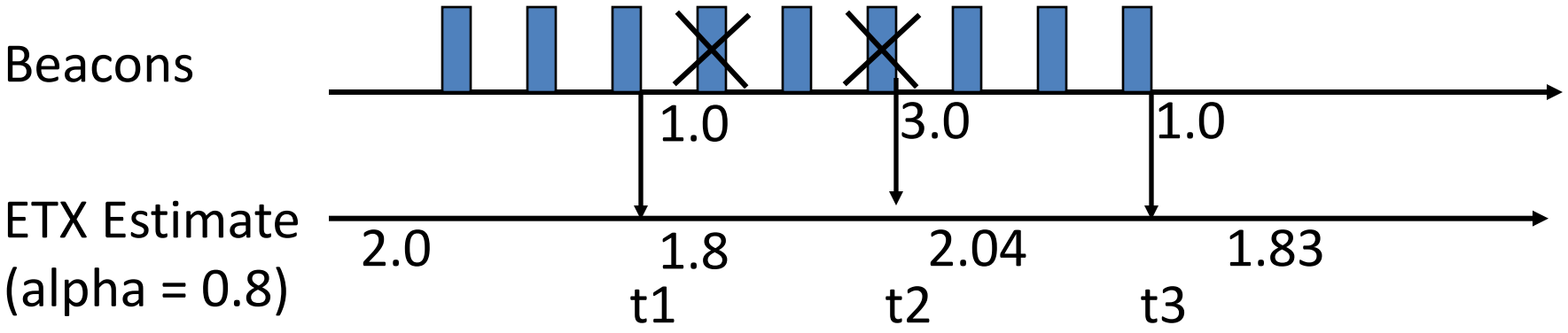
[from wikipedia]

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

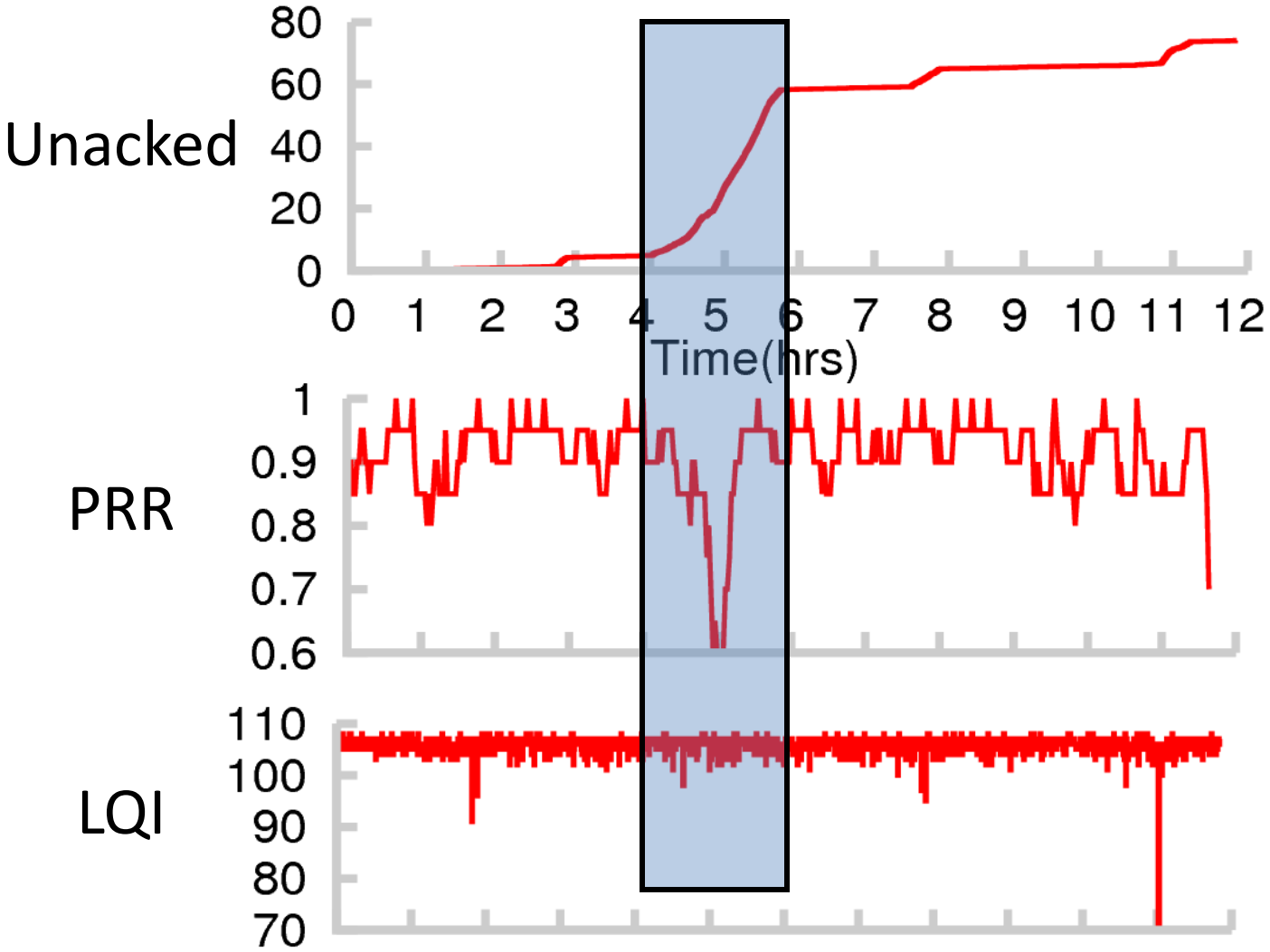
Link quality estimation

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

ETX Estimation Example



Link Estimation using PHY info

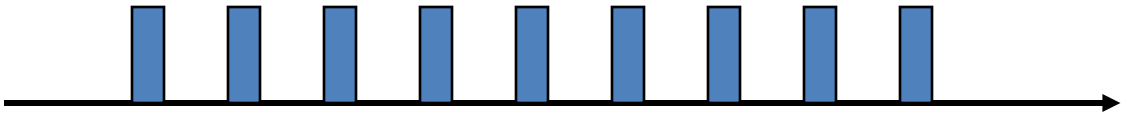


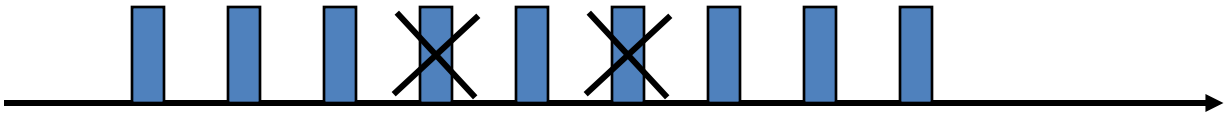
Quality of reception = Signal / Noise

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



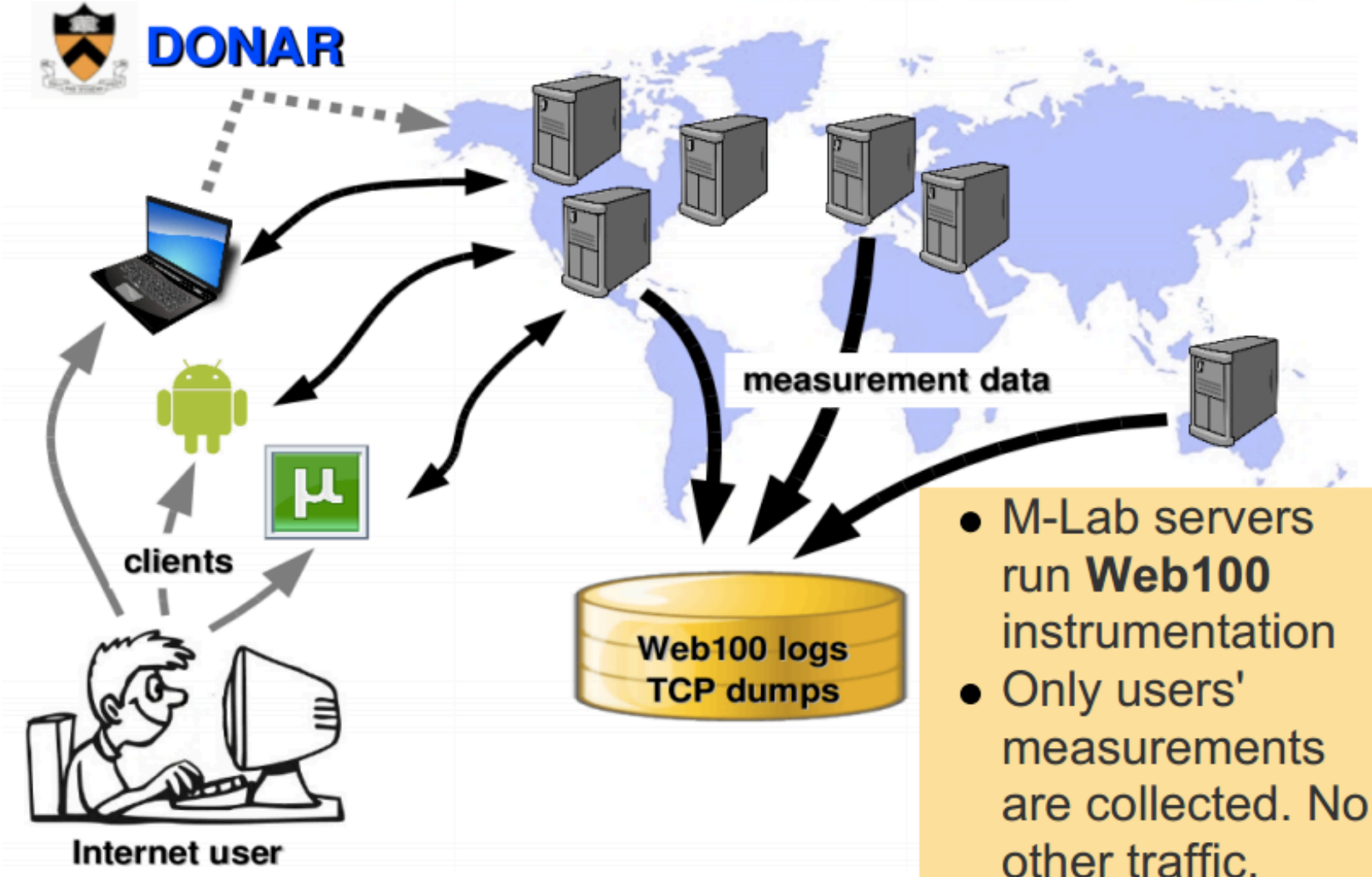


How to select participants for an HCI study?

What is the possibility of sampling bias?

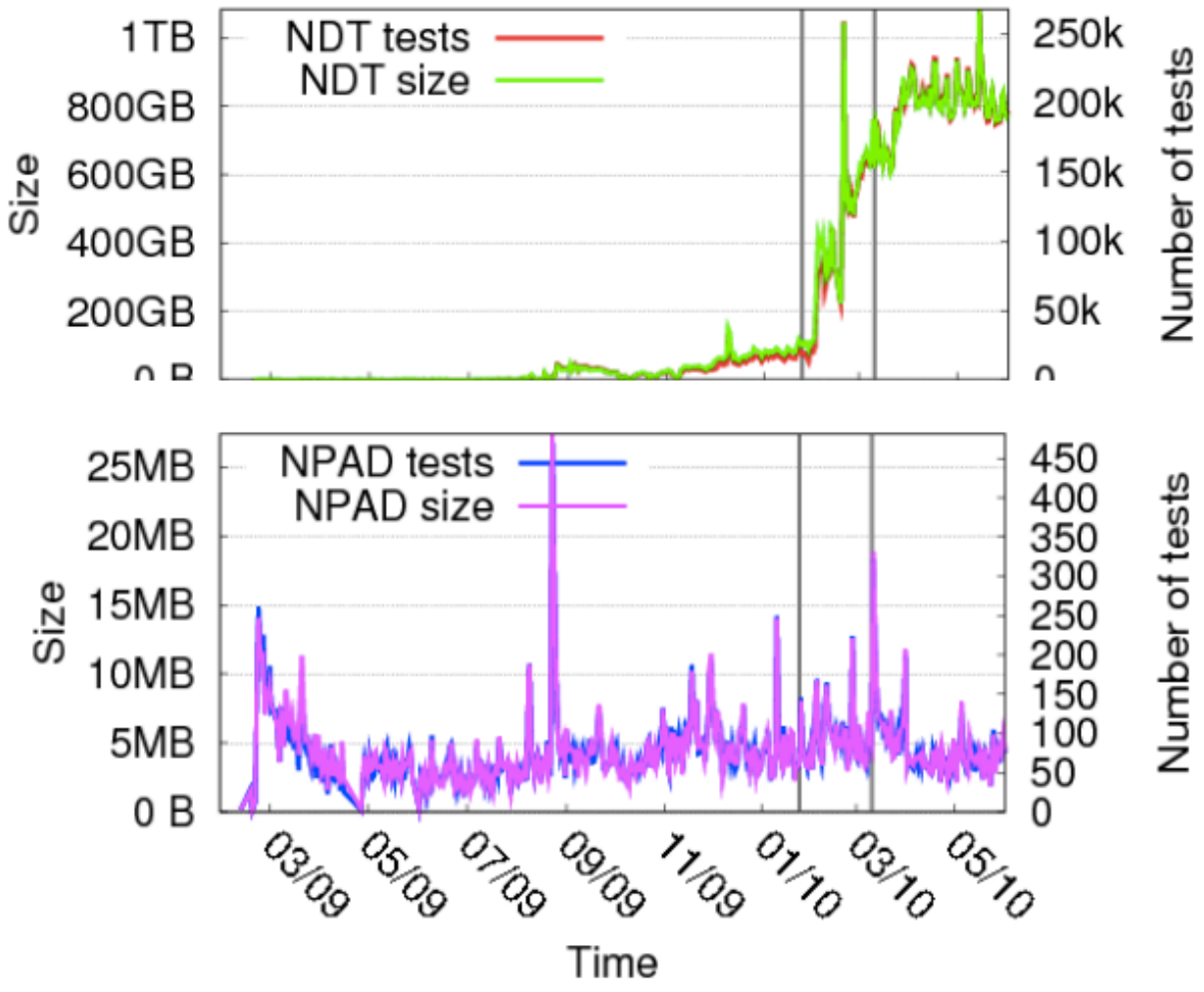
Example of bias due to a significant change in mixture of data source

Measurement & Data collection



- M-Lab servers run **Web100** instrumentation
- Only users' measurements are collected. No other traffic.

How much data? How many tests?



Jan 25 2010
uTorrent launch
Mar 11 2010
FCC launch

NDT

Tot tests: 22**M**

Tot size: 93**TB**

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?

Datasets and consequences

How Vector Space Mathematics Reveals the Hidden Sexism in Language

https://www.technologyreview.com/s/602025/how-vector-space-mathematics-reveals-the-hidden-sexism-in-language/

Exercise

Think of a big-data application

Identify a dataset on which you want to do "machine learning"

Describe potential bias there in the application

HW8 – Generating ideas

Pick a research paper.

Generate three ideas related to that paper.

- Title
- Similarity
- Difference