Research Methods in computer science

Spring 2017

Lecture 13

Omprakash Gnawali

March 1, 2017

Agenda

HW6 live grading

Research Conference Updates

Sampling/Dataset Bias

HW7

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.

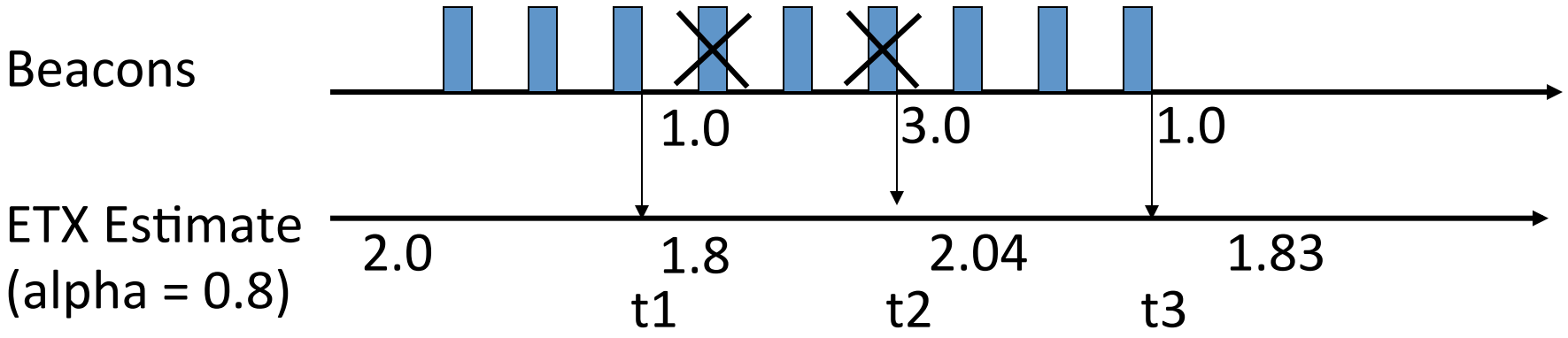
[from wikipedia]

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

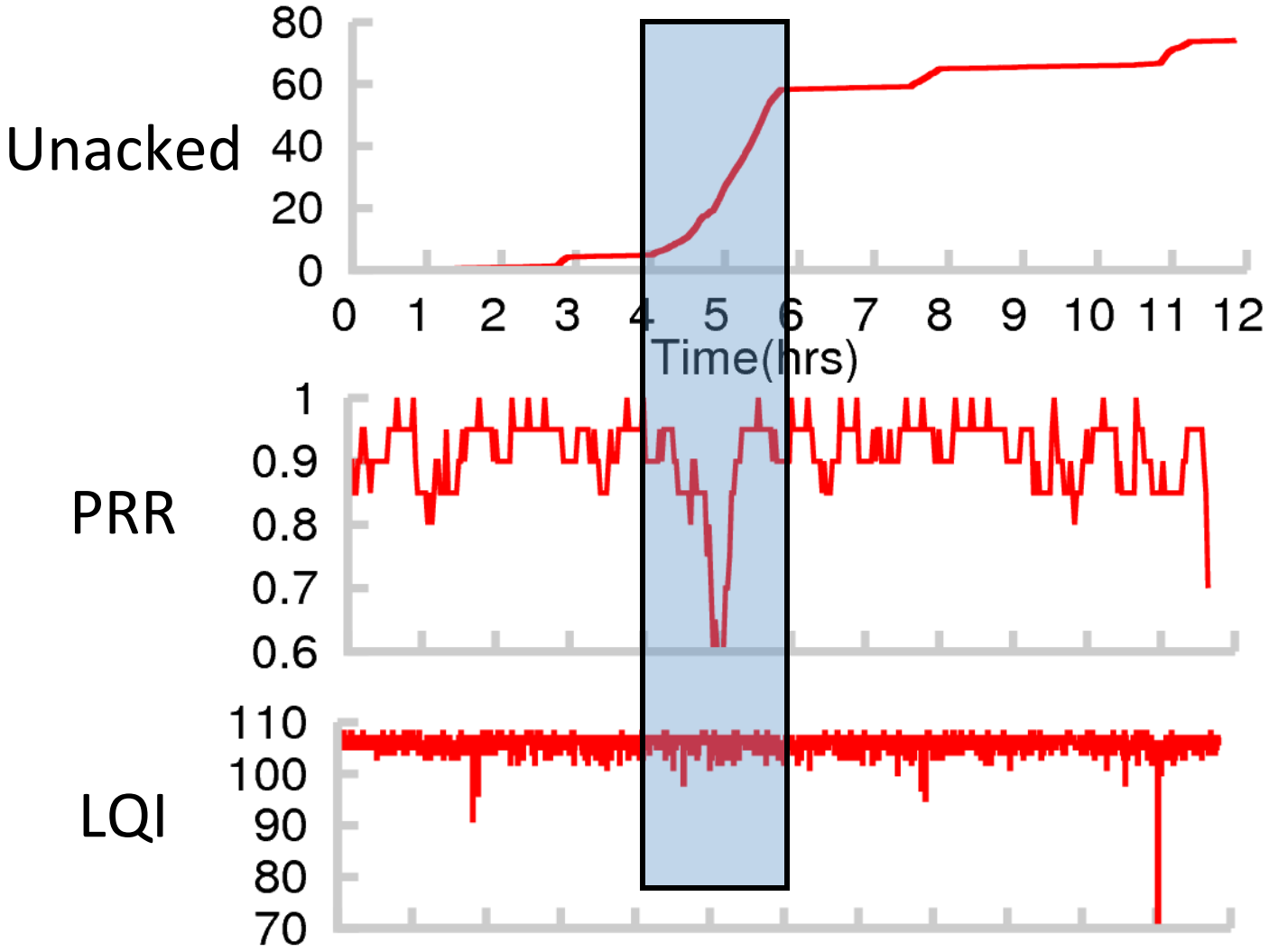
Link quality estimation

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

ETX Estimation Example



Link Estimation using PHY info

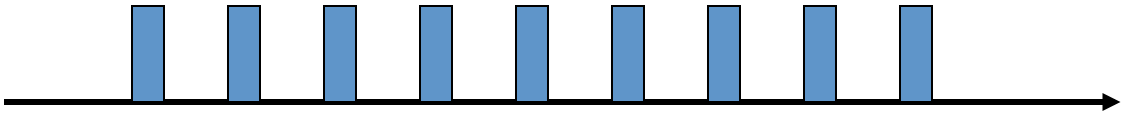


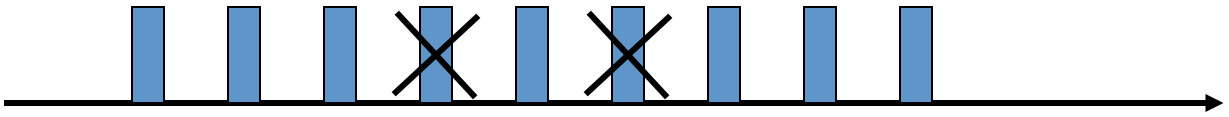
Quality of reception = Signal / Noise

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



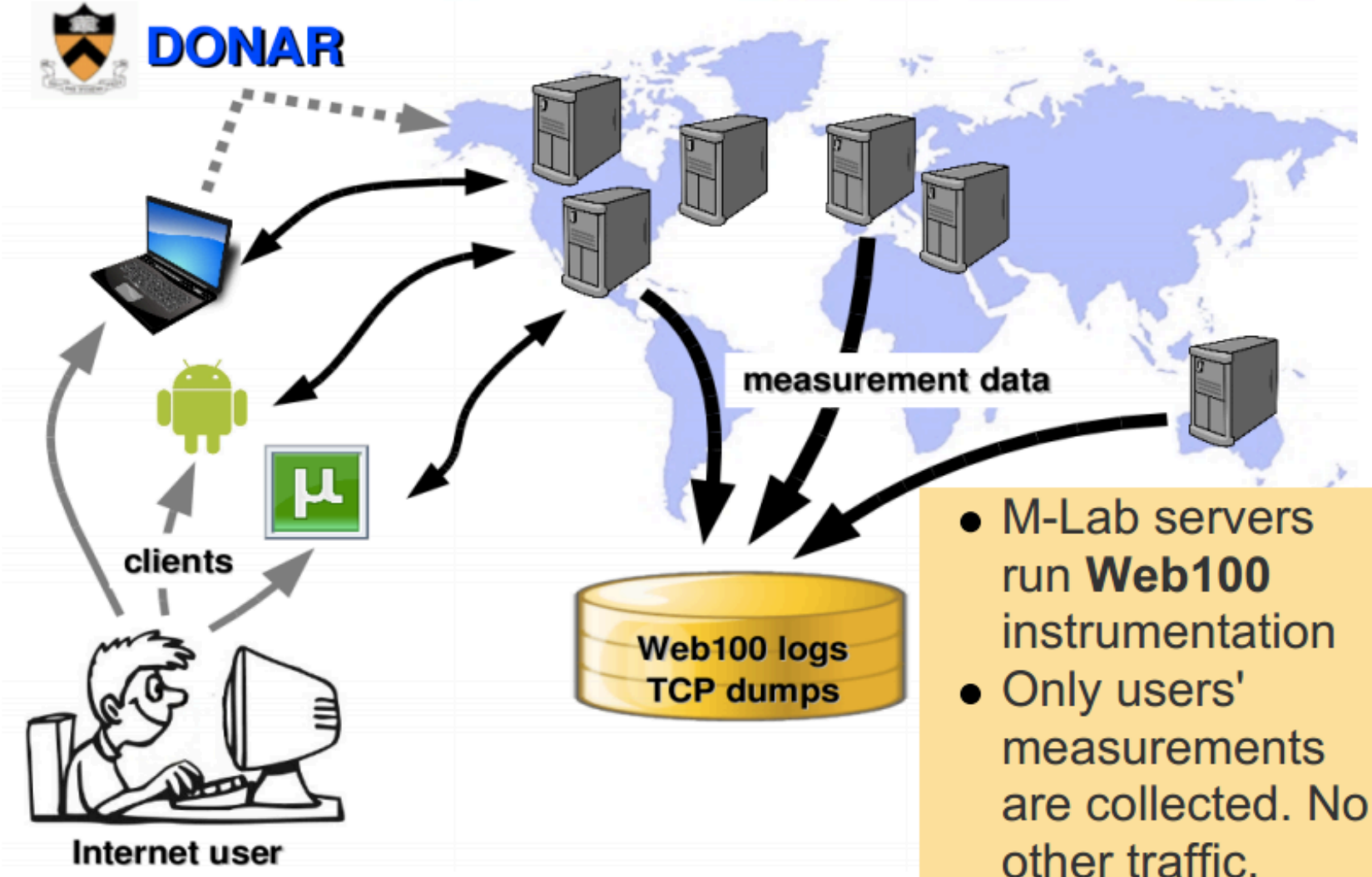


How to select participants for an HCI study?

What is the possibility of sampling bias?

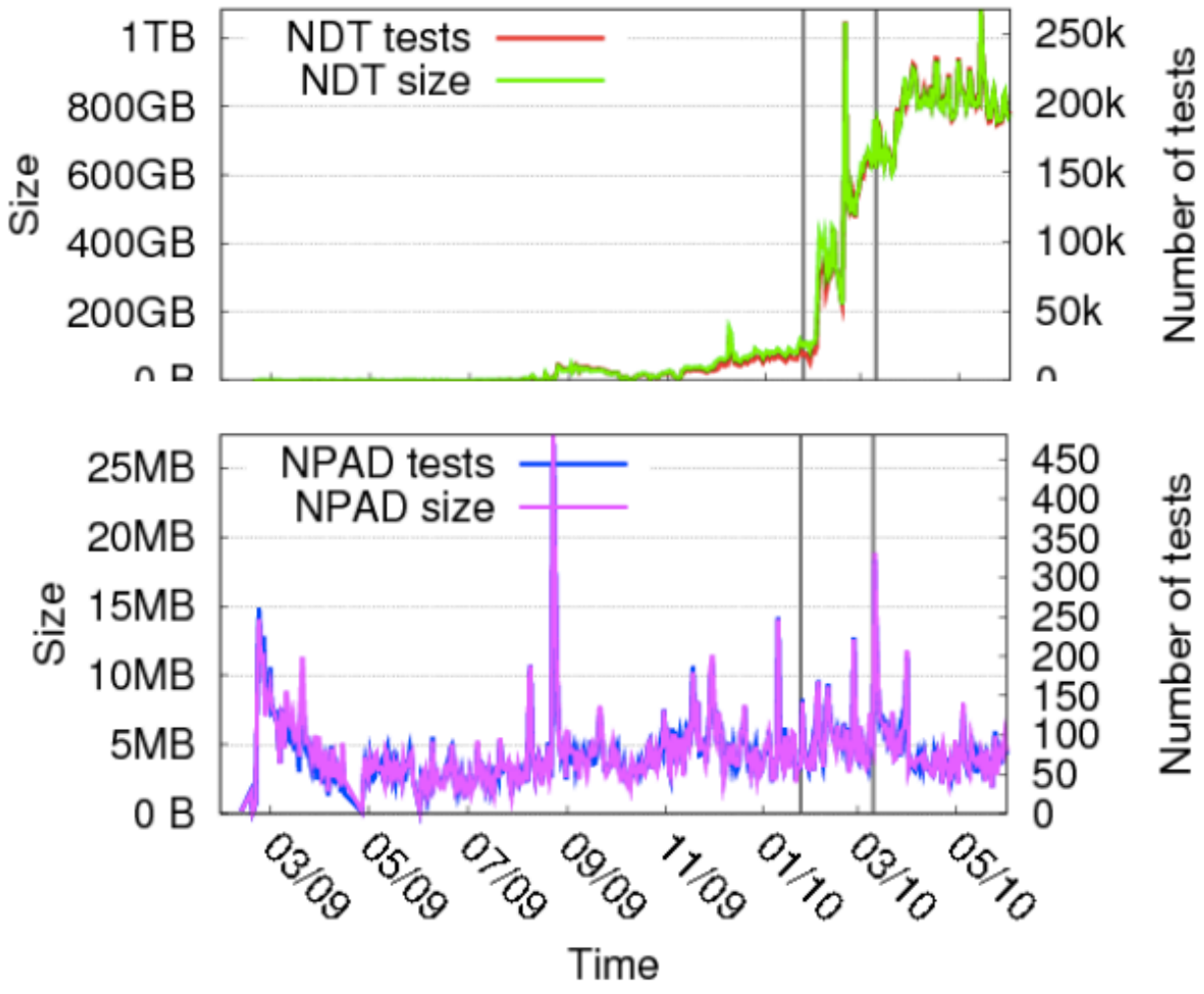
Example of bias due to a significant change in mixture of data source

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?

HW7 – Paper Introduction

Write the introduction section of your research paper. Please strictly follow the template we discussed for introduction. Copy-paste the questions [courtesy Widom] and write a paragraph below each question and complete your introduction.

What is the problem?

Why is it interesting and important?

Why is it hard? (E.g., why do naive approaches fail?)

Why hasn't it been solved before? (Or, what's wrong with previous proposed solutions? How does mine differ?)

What are the key components of my approach and results? Also include any specific limitations.

Summary of results and contributions.