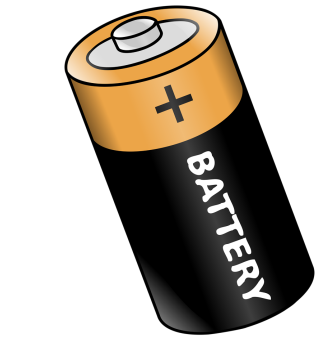
On Using Crowd-sourced Network Measurements for Performance Prediction

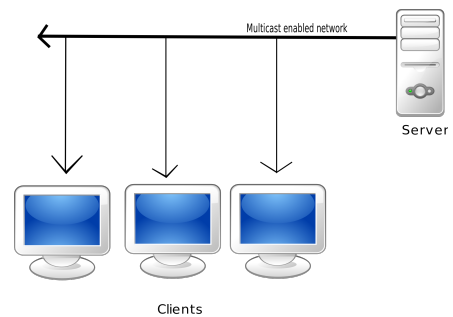
Pontus Persson, Tova Linder, Anton Forsberg,
Jakob Danielsson and Niklas Carlsson

LINKÖPING UNIVERSITY

Bandwidth predictions based on performance maps

- Minimize download times
- Reduce mobile units' energy usage
- Improve streaming efforts
- Efficient handovers in multi-homing environments





Contributions

1. Characterize the mobile speedtest usage of Bredbandskollen
2. Analyze the variation and predictability of download speeds within and between locations
3. Case-based performance analysis of different data sharing policies to build performance maps

1. **Characterize the mobile speedtest usage of Bredbandskollen**
2. Analyze the variation and predictability of download speeds within and between locations
3. Case-based performance analysis of different data sharing policies to build performance maps

LINKÖPING UNIVERSITY

Dataset and limitations

Dataset from Bredbandskollen, Sweden's primary internet speed test service maintained by the Internet Foundation in Sweden.
We have used measurements from:

- mobile (cellular) networks
- between Jan. 2014 to Feb. 2015

A grand total of 16 million measurements.

Dataset and limitations

Each measurement has many properties, including:

- GPS coordinates
- Upload/Download speeds
- Operator
- Technology (3G/4G)
- Time and date

Measurement concentration

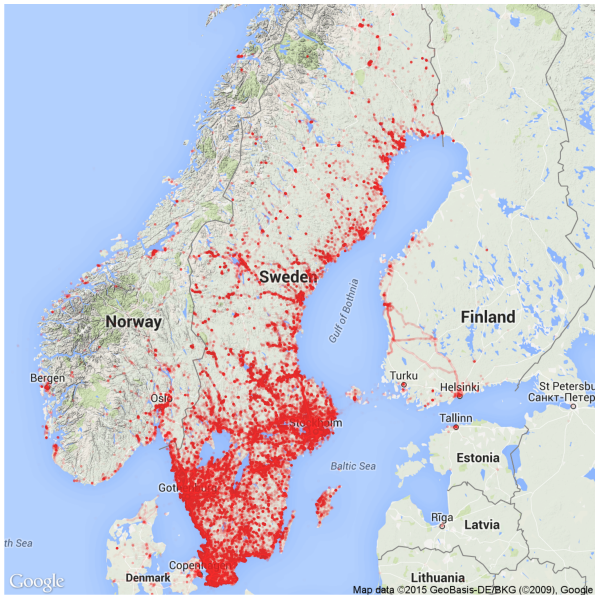
- Highly skewed towards densely populated areas
- Location sizes:

$$200 \times 200m^2$$

$$400 \times 400m^2$$

$$800 \times 800m^2$$

$$1600 \times 1600m^2$$



Measurement concentration

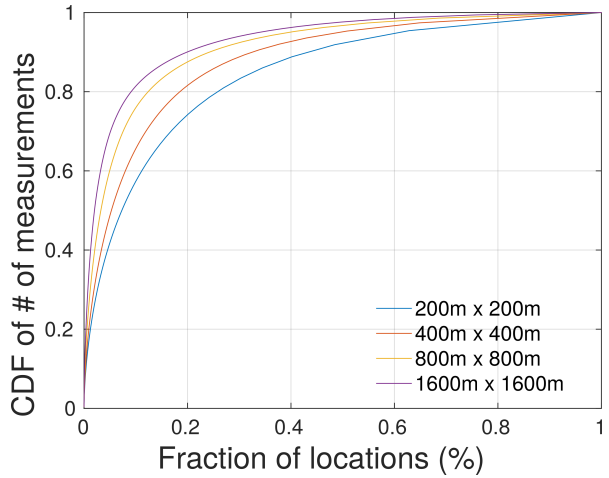
- Highly skewed towards densely populated areas
- Location sizes:

$$200 \times 200m^2$$

$$400 \times 400m^2$$

$$800 \times 800m^2$$

$$1600 \times 1600m^2$$



Measurement concentration

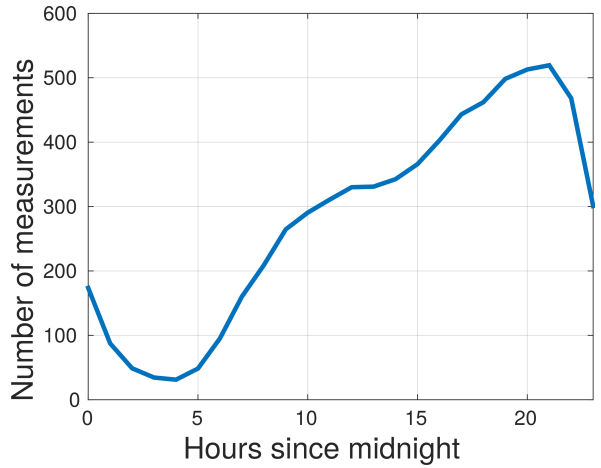
- Highly skewed towards densely populated areas
- Location sizes:

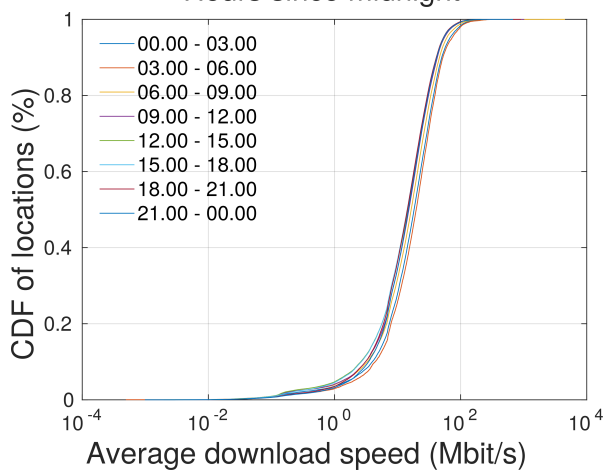
$$200 \times 200m^2$$
$$400 \times 400m^2$$
$$800 \times 800m^2$$
$$1600 \times 1600m^2$$

- Thresholds $15, 20, 25, 30$ for the location sizes respectively
- At least 15% of all locations
- At least 70% of all measurements

Time-of-day analysis

- Clear diurnal pattern
- Biggest difference between
  - 3:00-6:00 (20.7Mbit/s)
  - 18:00-21:00 (18.2Mbit/s)





1 Characterize the mobile speedtest usage of Bredbandskollen

2 Analyze the variation and predictability of download speeds within and between locations

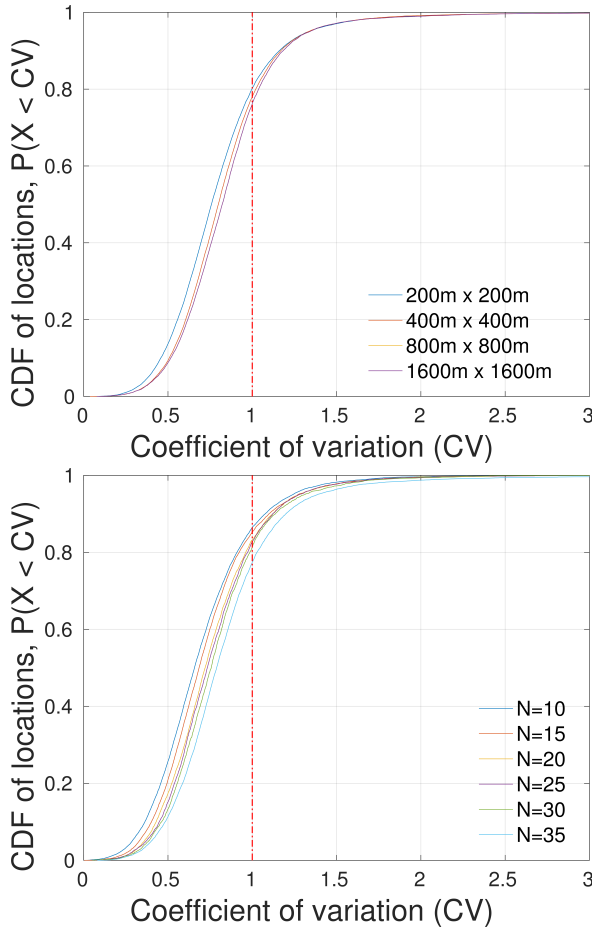
3 Case-based performance analysis of different data sharing policies to build performance maps

LINKÖPING UNIVERSITY

Single-location Variation Analysis

Relative variation $CV = \frac{\sigma_i}{\mu_i}$ for location i.

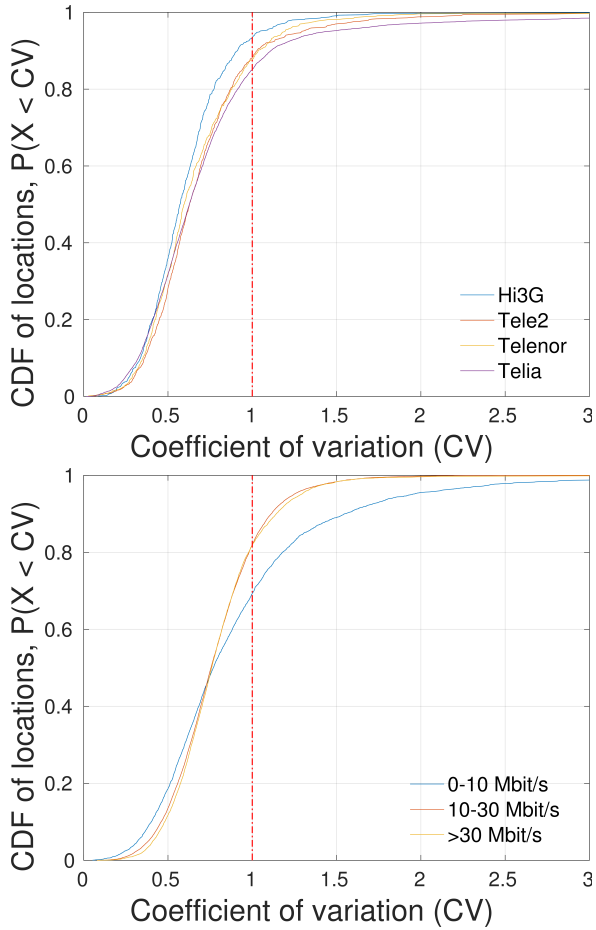
- $CV < 1 \Rightarrow$ easier to predict
- Tested by varying:
  - Location size
  - Measurement threshold
  - Operator
  - Average download speed



Single-location Variation Analysis

Relative variation $CV = \frac{\sigma_i}{\mu_i}$ for location i.

- $CV < 1 \Rightarrow$ easier to predict
- Tested by varying:
  - Location size
  - Measurement threshold
  - Operator
  - Average download speed



Pairwise Head-to-Head Comparison

When one operator provides better download speeds in a location, knowing the “winner” is fortunate for:

- users using multi-homing
- multipath-TCP users

Difference in average download speed between operators tested using Welch’s t-test.

Pairwise Head-to-Head Comparison

Percentage of locations where average download speeds differ with 95% confidence.

	Telia	Tele2	Telenor	Hi3G
Telia	-	50.0%	45.5%	46.7%
Tele2	50.0%	-	37.3%	63.5%
Telenor	45.5%	37.3%	-	42.9%
Hi3G	46.7%	63.5%	42.9%	-
All	14.5%	25.8%	26.1%	25.1%

Comparing Neighbor Locations

Geographic download speed differences allow mobile client to select opportune times and locations to download content to:

- improve download speeds
- save energy
- improve conditions for delay sensitive applications

Difference in average download speed between neighbouring locations tested using Welch's t-test.

Comparing Neighbor Locations

With at least 20 measurements in each location and 95% confidence there is a "winner" in:

- 42.2% of $200 \times 200m^2$ location pairs
- 47.5% of $400 \times 400m^2$ location pairs
- 53.9% of $800 \times 800m^2$ location pairs
- 57.5% of $1600 \times 1600m^2$ location pairs

Locations with more measurements or high average speed are more often in a pair where there's a "winner".

1 Characterize the mobile speedtest usage of Bredbandskollen

2 Analyze the variation and predictability of download speeds within and between locations

3 Case-based performance analysis of different data sharing policies to build performance maps

LINKÖPING UNIVERSITY

Simulations

Simulates user moving along a path of N locations predicting download speeds using policy specified measurements.

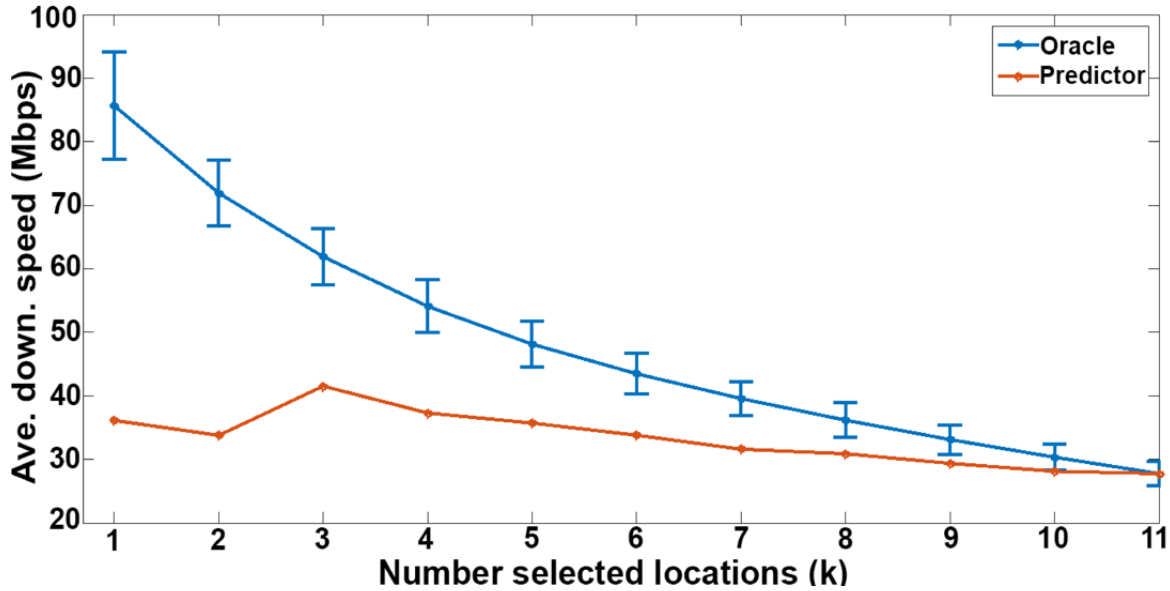
- $1km^2$ locations
- Measurements from the Stockholm area
- User download in the k best locations and is compared to an oracle

Policies

- **Full sharing:** No restrictions imposed on what measurements to use
- **Same operator:** Only measurements from the same operator used
- **Same technology:** Only measurements with the same technology (3G/4G) are used
- **Restricted sharing:** A combination of *Same operator* and *Same technology*
- **Random sharing:** For each locations, users only use information of $p\%$ of randomly selected measurements

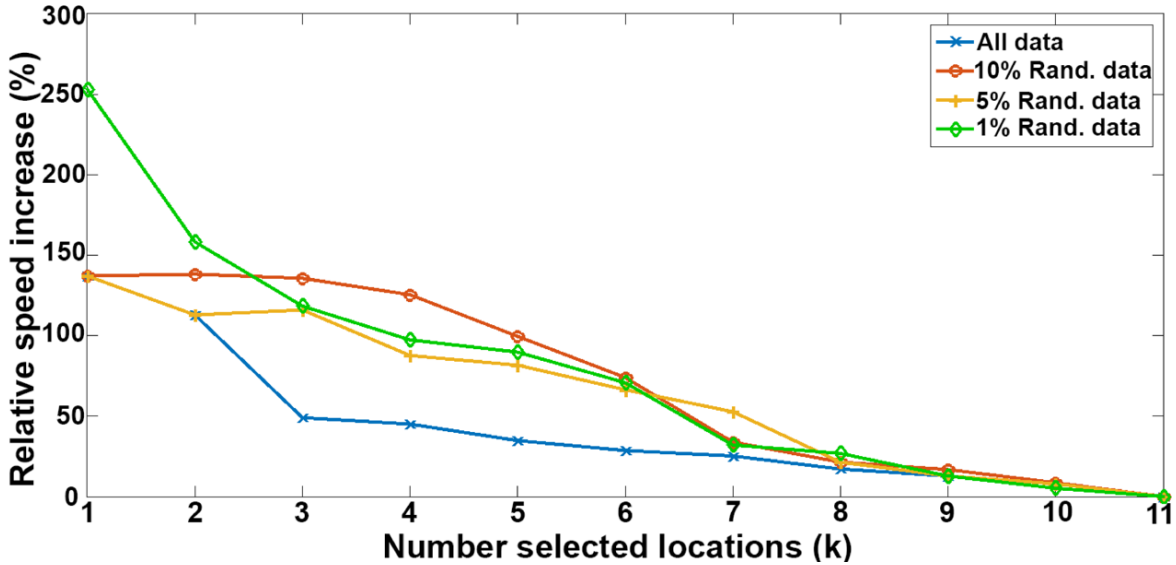
Full sharing

- Room for improvement
- Much better than “No sharing”



Random sharing

- Full sharing consistently better
- Less information ⇒ worse prediction



Policy comparison

- Our selective policies perform best
- "Same technology" and "Restricted sharing" are the two winners

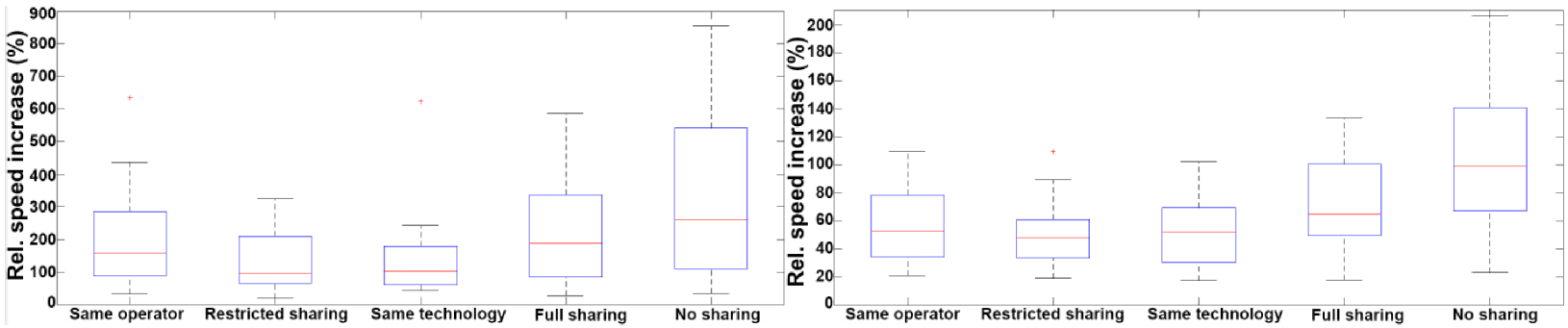


Figure: $k = 1, N = 11$ and $k = 4, N = 11$

Policy comparison

- Now “Full sharing” outperforms the other policies.
- More important with sufficient number of measurements

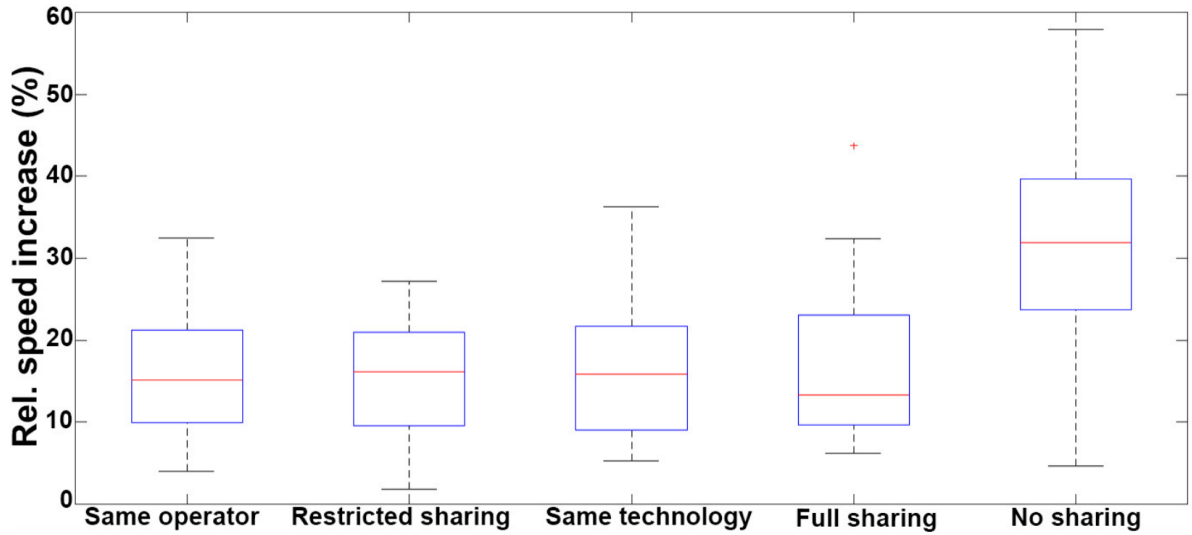


Figure: $k = 8, N = 11$

Conclusions

- Beneficial to make predictions in large locations
- Locations with high average download speeds are easier to predict
- Multi-homing and multi-path TCP can be beneficial
- Prediction accuracy increases with filtered data on technology and operator when downloading at a few locations

On Using Crowd-sourced Network Measurements for Performance Prediction

Pontus Persson, Tova Linder, Anton Forsberg,
Jakob Danielsson and Niklas Carlsson

www.liu.se

LINKÖPING UNIVERSITY