RICERCANDO:
Towards Mobile Broadband Measurement Mining

Veljko Pejović

Faculty of Computer and Information Science
(FRI)
University of Ljubljana, Slovenia

Veljko.Pejovic@fri.uni-lj.si

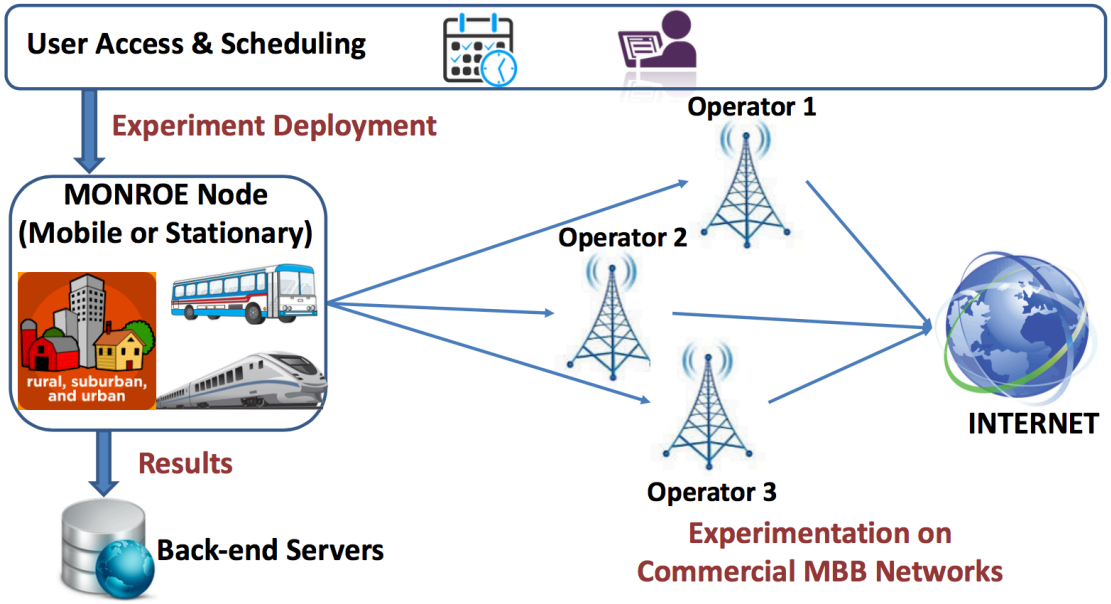
University *of Ljubljana*
Faculty *of Computer and*
Information Science

Cyber Alp Retreat,
Gstaad, Switzerland, January 2018

Fabio Ricciato sends his greetings to everyone!

MONROE

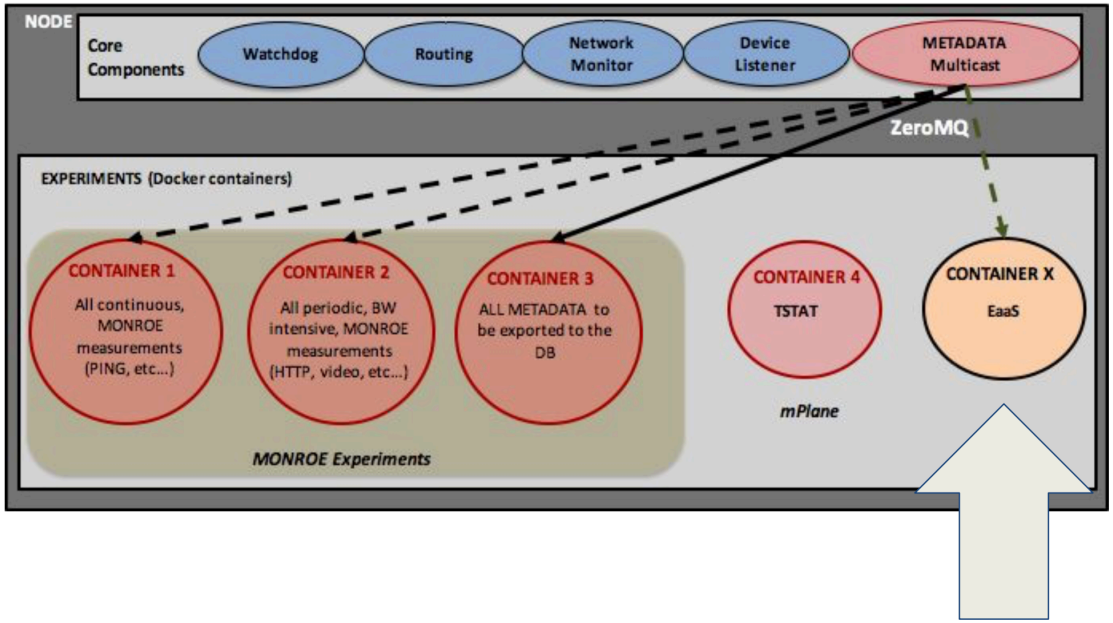
Design, build and operate an open, European-scale, and flexible hardware-based platform to run experiments on operational 3G/4G Mobile Broadband networks with WiFi connectivity



Experience: An Open Platform for Experimentation with Commercial Mobile Broadband Networks, Alay et al., ACM MobiCom'17

MONROE

- Experimentation:
  - Users create Docker containers with experiments that are uploaded to individual nodes and ran at scheduled times
  - Background ping throughput, traceroute, etc. measurement ran periodically



Network Measurement Data

- Performance indicators:
  - Packet delay, jitter, packet loss, open/closed ports, etc. whatever users design
- Metadata (in MONROE):
  - Ping time, upload/download speed
  - Network type (2G, 3G, LTE, WiFi)
  - Signal strength (RSSI, RSRP)
  - Node coordinates
  - Node CPU utilisation, temperature
  - Other experiments on a node, software versions, device config

With crowdsourced measurements also:

- Device model, OS
- Device artefacts (e.g. bent antenna)
- User's mobile plan and potential caps reached

RICERCANDO

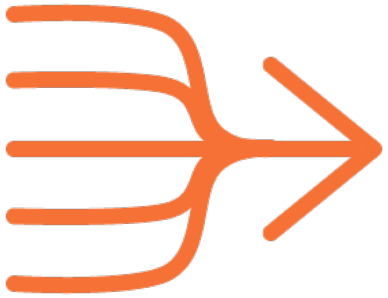
- Goal: Facilitate network measurement data (and metadata) exploration, visualisation and interpretation
  - Ease problem discovery and troubleshooting
  - Separate monitoring system glitches from network anomalies
  - Support Root-Cause Analysis (RCA)

RICERCANDO Approach

- Develop a visual programming environment for data analysis
  - Jupyter-based rapid data exploration
  - Orange-based
    data analytics pipeline
    - "Data mining fruitful and fun" tool developed at FRI
    - Used in bioinformatics, statistics offices, etc.
    - http://orange.biolab.si/

Challenges

- Big time-series data
  - A few GB of data per day
  - Events with up to 10ms granularity
- Data merging
  - Different data comes with different granularity
  - Mismatched timestamps among different data sources
- Demand for visually-driven analysis
- Identify the most useful analytics tools

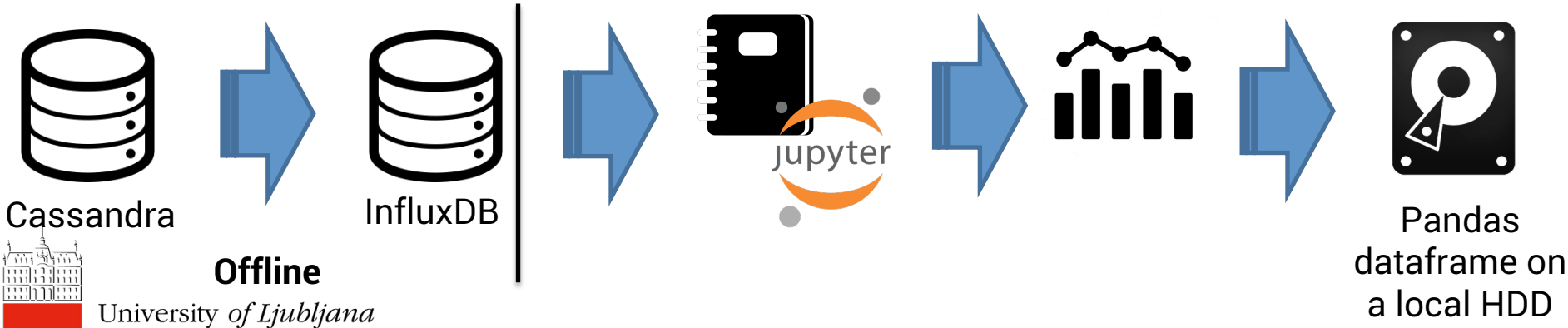


Temporal data handling and merging

- Standard databases are insufficient
  - From Cassandra to Influx DB
    - From noSQL to a DB optimised for time series data
    - “up to 168x faster response times for tested queries”
- Data merging:
  - Create 10ms granularity tables, but support user-defined granularity querying
  - Allow users to query the DB and get results interpolated in the desired way:
    - Without interpolation or with any of the pandas supported methods (e.g. linear, nearest, etc.)

Visual analysis

- Jupyter notebooks
  - Orange is very easy to use but the overhead prevents the analysis of massive data sets
  - Jupyter Notebooks allow highly customised analysis with substantial machine learning and visualisation support
  - Pipeline:

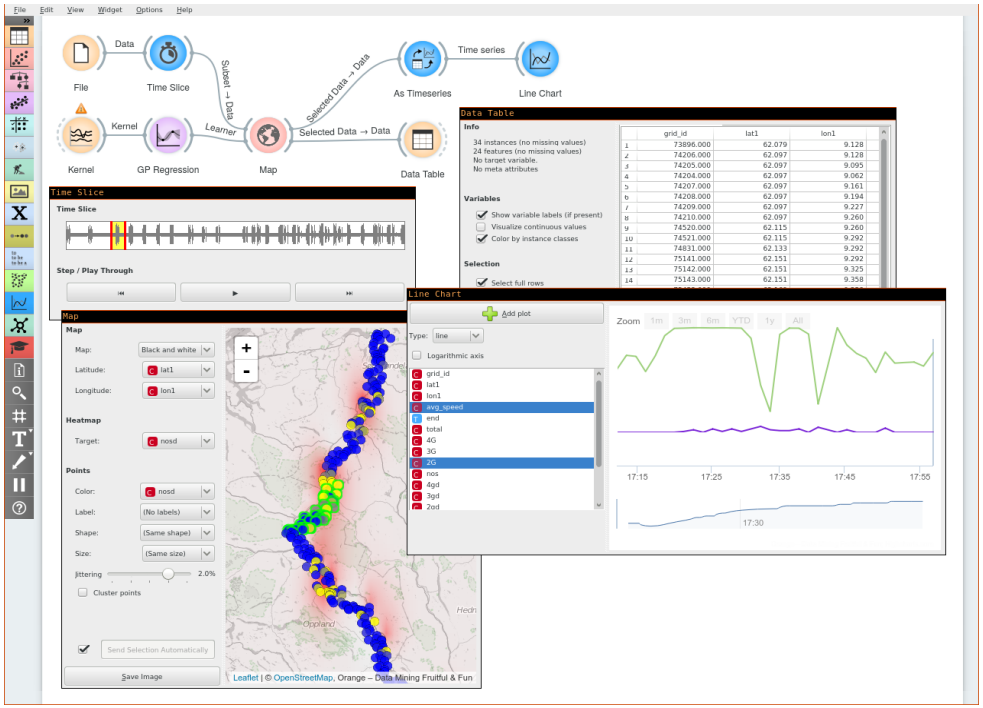


Visual analysis

- Orange analysis
  - Once we single out interesting data we send it for sophisticated analysis in Orange
  - Pipeline

Identifying the Most Useful Tools

- Geo-referenced visualisation:
  - Signal coverage maps
  - Statistical comparison of measurements in different regions
  - Location-dependent anomaly detection

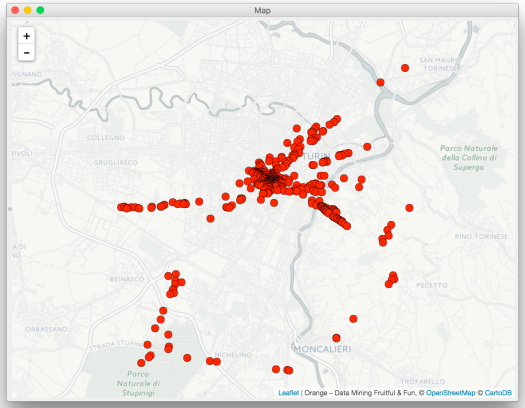


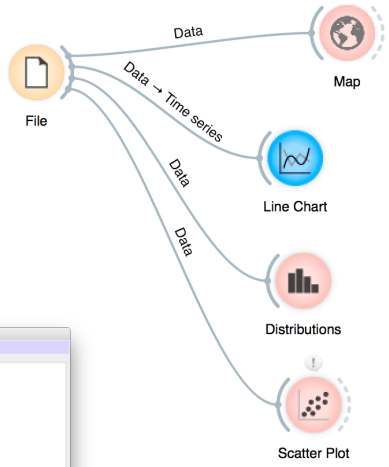
Identifying the Most Useful Tools

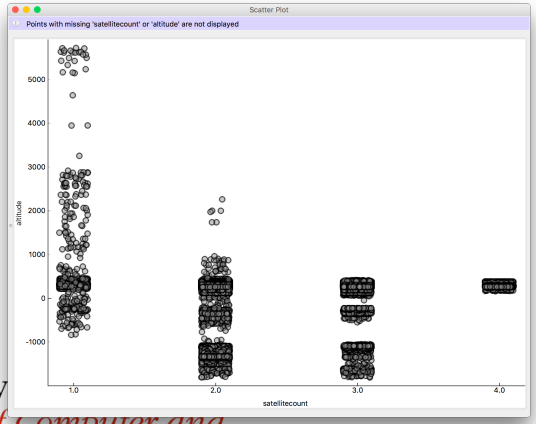
- Significant factor analysis
  - Which (combinations of) attributes impact(!) measurements results and how much?
  - For different types of data we implement different tests:
    - Hypergeometric test example
      - *N* – population size (say, all port 25 blocking tests)
      - *K* – success count
      - *n* – number of draws (say, all tests with iPhone on LTE)
      - *k* – number of observed successes
      - *Enrichment = (k/n)/(K/N)*

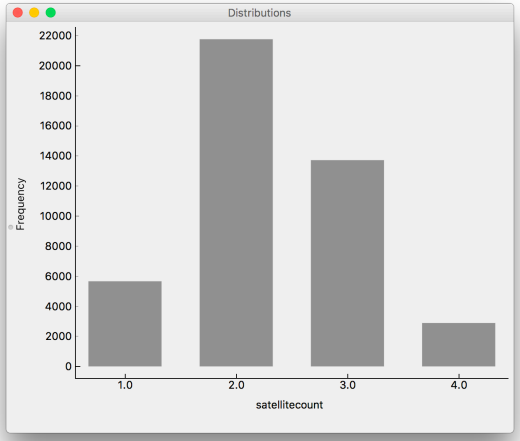
RICERCANDO in Action

- Anomaly examination



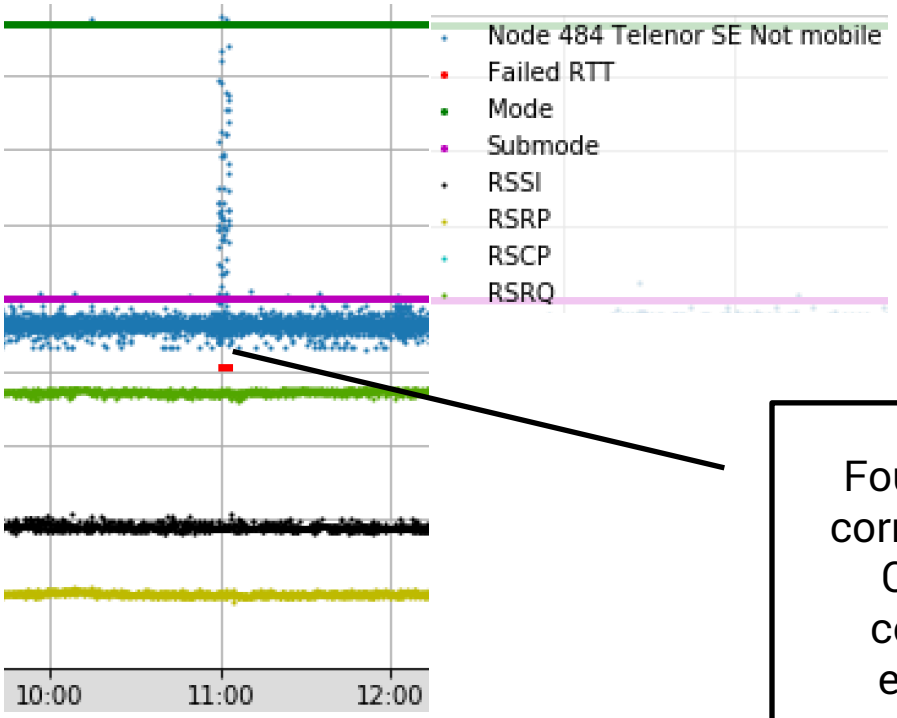






RICERCANDO in Action

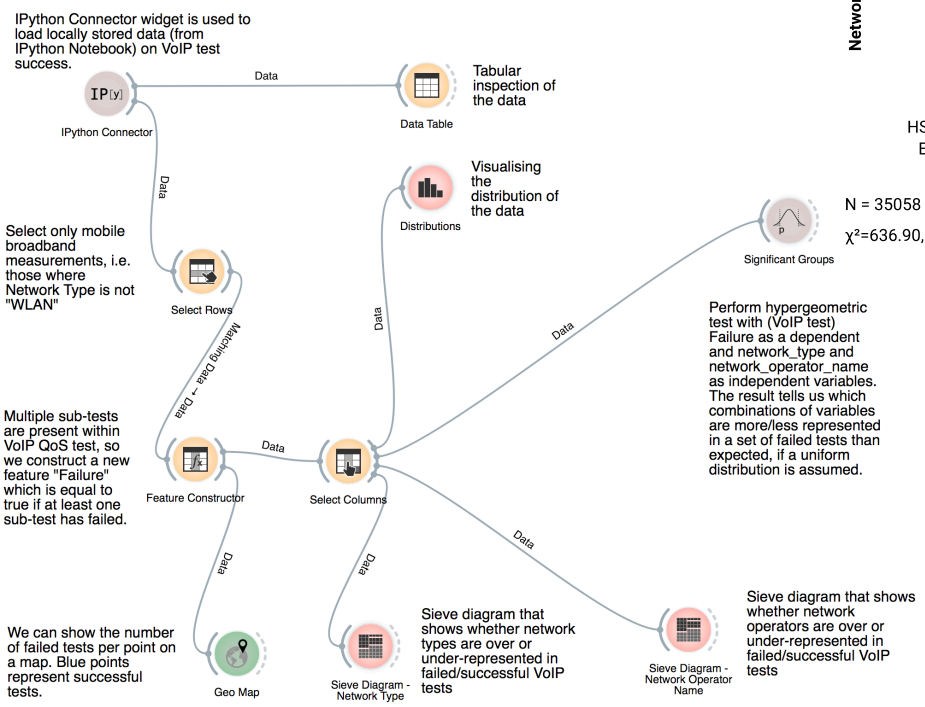
- Experiment interplay – visual analysis

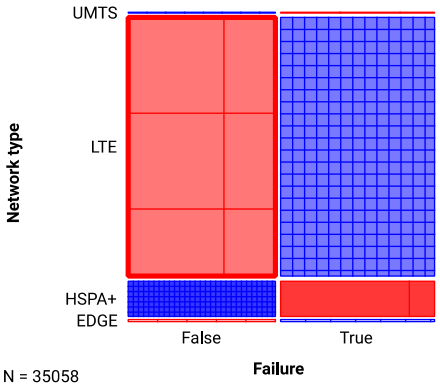


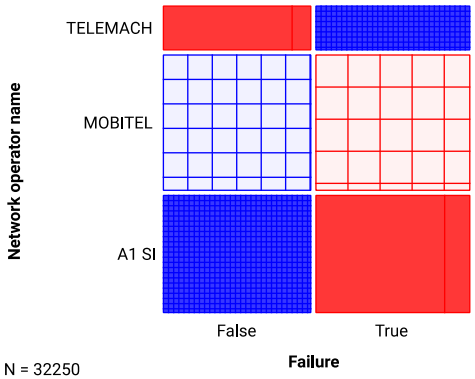
Found that RTT spikes correlate with increased CPU activity which correlates with user experiment activity

RICERCANDO in Action

- Analysing VoIP test performance for Slovenian mobile broadband operators







RICERCANDO

- A toolbox for supporting discovery in mobile broadband data
- Orange tools are publicly available:
  - Main environment http://orange.biolab.si/
  - Add-on enabling local InfluxDB mining and stored Jupyter data access with additional tests https://github.com/biolab/monroe-anal/
- Mine your data, not Bitcoins!

…since it’s my first time at Cyber Retreat

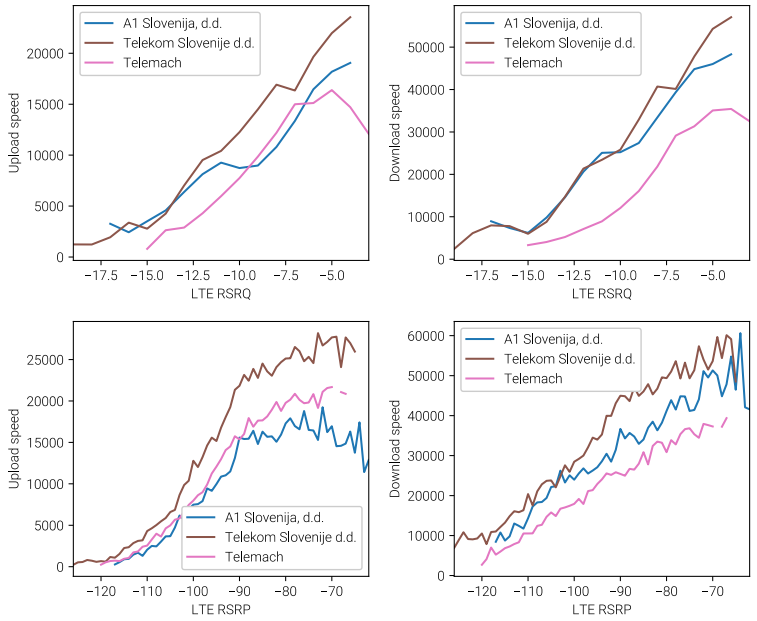
University *of Ljubljana*
Faculty *of Computer and Information Science*

Mobile and Wireless Research at FRI

Mobile broadband measurement data analysis

Collaborative wireless decoding

Mobile sensing for behaviour modelling

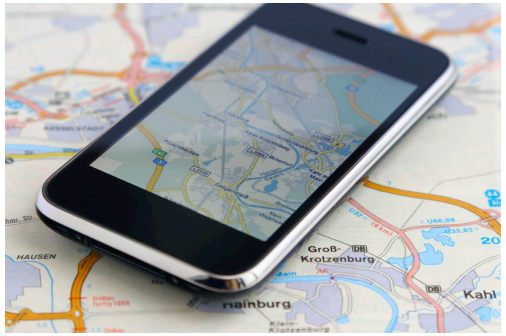


University of Ljubljana
Faculty of Computer and Information Science

Mobile Sensing for Behaviour Modelling Projects

Big Data

- Telecom call data records
- Modelling and examining disease spreading and containment
  - Where do people move – How does the disease spread
  - Who do people call –How does the information spread?



Smaller Data

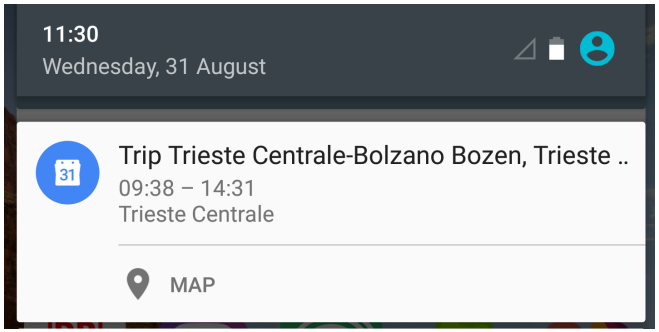
- Smartphones are personalised devices carried at all times
- Smartphone sensors readings can reveal (semantic) location, physical activity, communication patterns, emotion, etc.

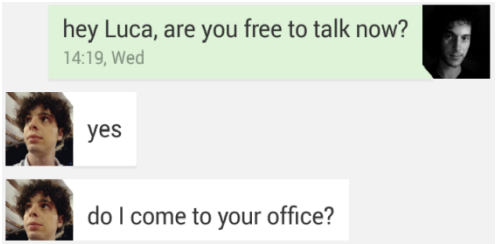
University *of Ljubljana*
Faculty *of Computer and Information Science*

A. Lima, M. De Domenico, V. Pejovic and M. Musolesi Disease Containment Strategies based on Mobility and Information Dissemination Nature Scientific Reports (2015).

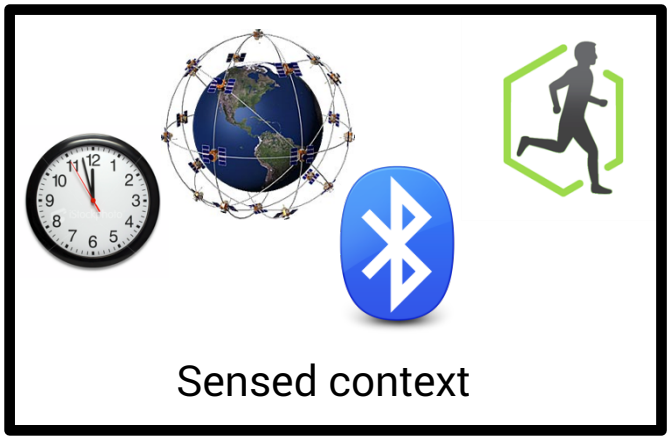
Mobile Sensing for Behaviour Modelling Project Example

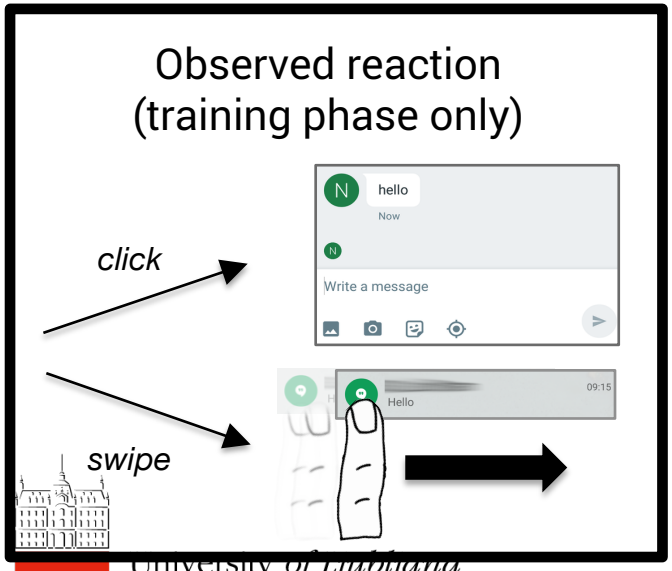
- Managing attention of a mobile user
  - When/how to efficiently deliver information to a user?





Mobile Sensing for Behaviour Modelling Project Example





Classifiers:
- Batch, online
- Naïve Bayesian, Tree-based, Boosting

Notification response presence, time, sentiment

InterruptMe Android Library for notification management

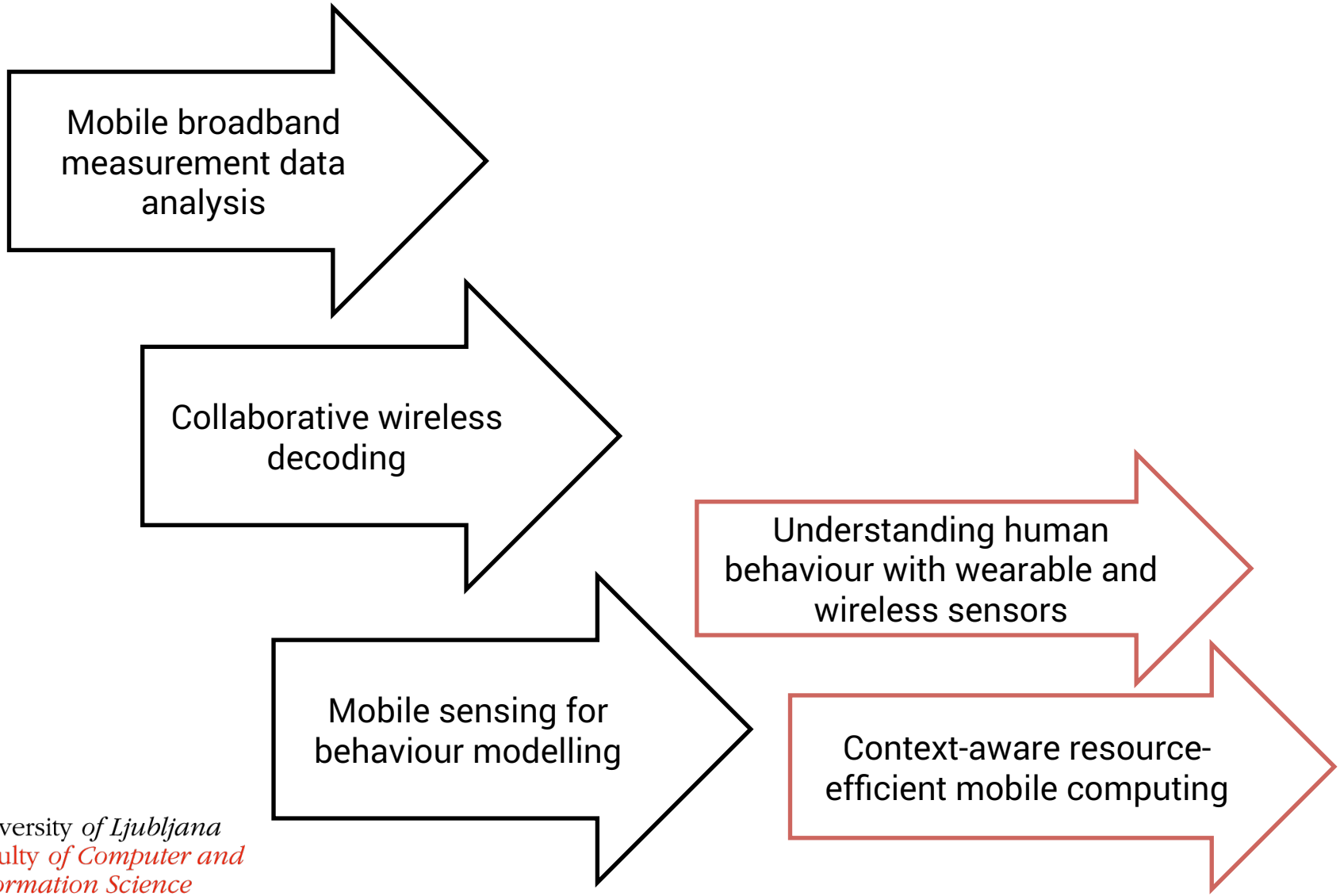
bitbucket.org/veljkop/intelligenttrigger

github.com/vpejovic/MachineLearningToolkit

Mobile Sensing for Behaviour Modelling Project Example

- Managing attention of a mobile user
  - When/how to efficiently deliver information to a user?
- Our attempts:
  - Using sensor data and experience sampling, on-device machine learning
  - Monitoring real-world notification usage
  - Taking into account application type, contact type (e.g. family, friend, work-related)
  - Sensing cognitive engagement with a phone

Mobile and Wireless Research at FRI



University *of Ljubljana*
Faculty *of Computer and Information Science*

Collaborators

- FRI:
  - Prof Blaž Zupan
  - Blaž Repas
  - Ivan Majhen
  - Dr Miha Janež
- Outside:
  - Prof Mirco Musolesi, Dr Abhinav Mehrotra (UCL)
  - Christoph Anderson (Kassel University)
- In between:
  - Prof Fabio Ricciato

Thank you!

Veljko Pejović
FRI, University of Ljubljana

Veljko.Pejovic@fri.uni-lj.si
http://lrss.fri.uni-lj.si/Veljko
@veljkoveljko

We have an opening for a professor in cybersecurity closing on 20/2!

University *of Ljubljana*
Faculty *of Computer and Information Science*