

# Real Time Movement Labeling of Mobile Event Streams

Elis Kõivumägi <sup>1,2</sup>, Mati Vait<sup>2</sup>, Amnir Hadachi<sup>1</sup>, Georg Singer<sup>2</sup>

<sup>(1)</sup>University of Tartu, Institute of Computer Science, Distributed Systems Group  
J.Liivi 2-311, Tartu 50409, Estonia

<sup>(2)</sup>Demograaft Project, Software Technology and Applications Competence Center, Ülikool 2,  
Tartu 51003, Estonia

# Agenda

- Background
- Nature of data
  - Stream
  - CDR
  - Cellplan
  - Test group
- Location detection
- Real time labeling
- Experiment
- Conclusion and future work
- Demo

# Background

- Involved parties
  - STACC
  - Regio/Reach-U
  - Tartu University
  - Positium
- Financing
  - Regio/Reach-U
  - EU

STACC Software Technology and  
Applications Competence Center

Reach  U

 UNIVERSITY OF TARTU  
1632

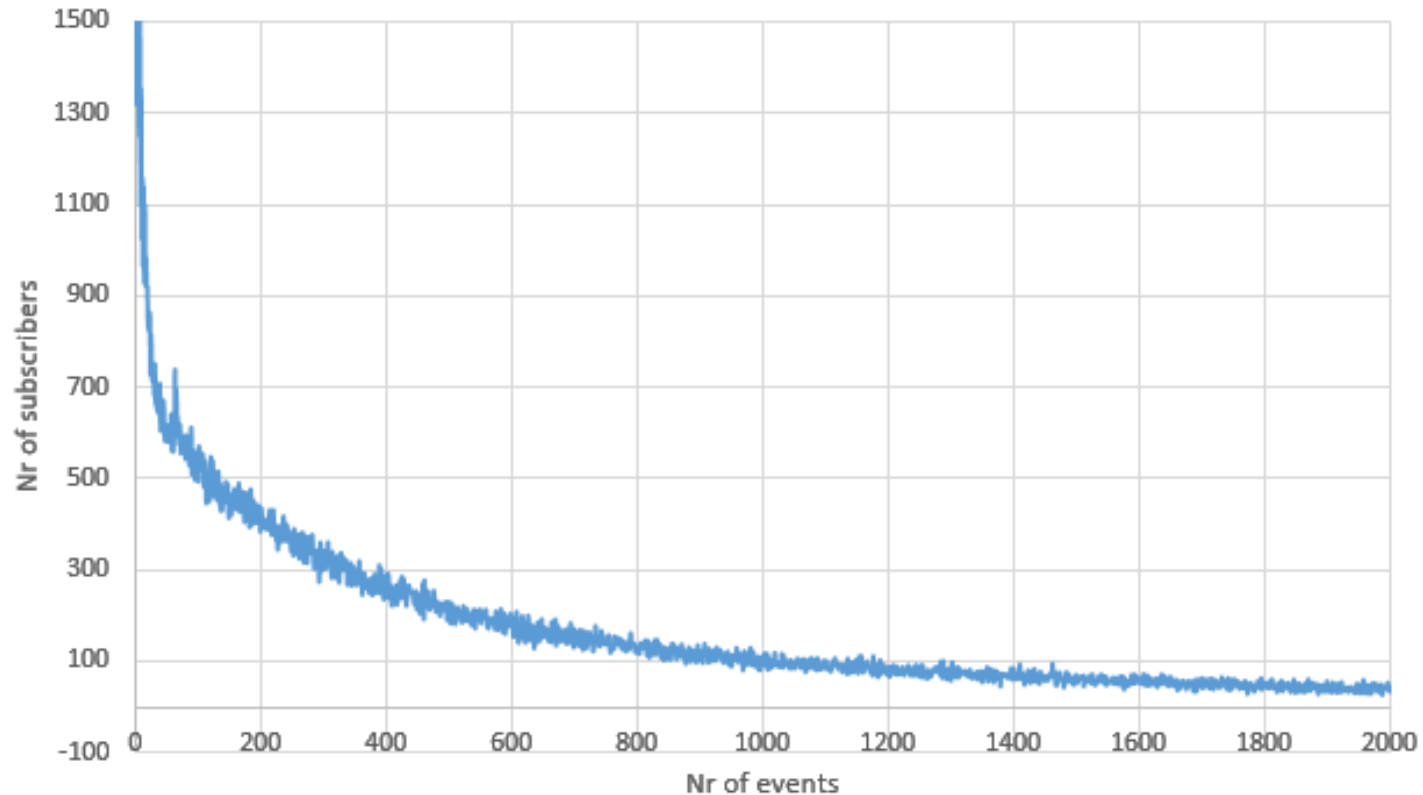
 positium

# Nature of data – whole dataset

- Averages per month over 3 months
- Stream
  - Subscriber count: ~400K
  - Event count: ~330M
  - Avg # events per subs: ~825 (without data events)
  - Median: 440
  - Lower quartile: 150
  - Higher quartile: 1112
- CDR
  - Subscriber count: ~700K
  - Event count: ~450M
  - Avg # event per subs: ~640 (with data events)
  - Median: 71
  - Lower quartile: 8
  - Higher quartile: 275

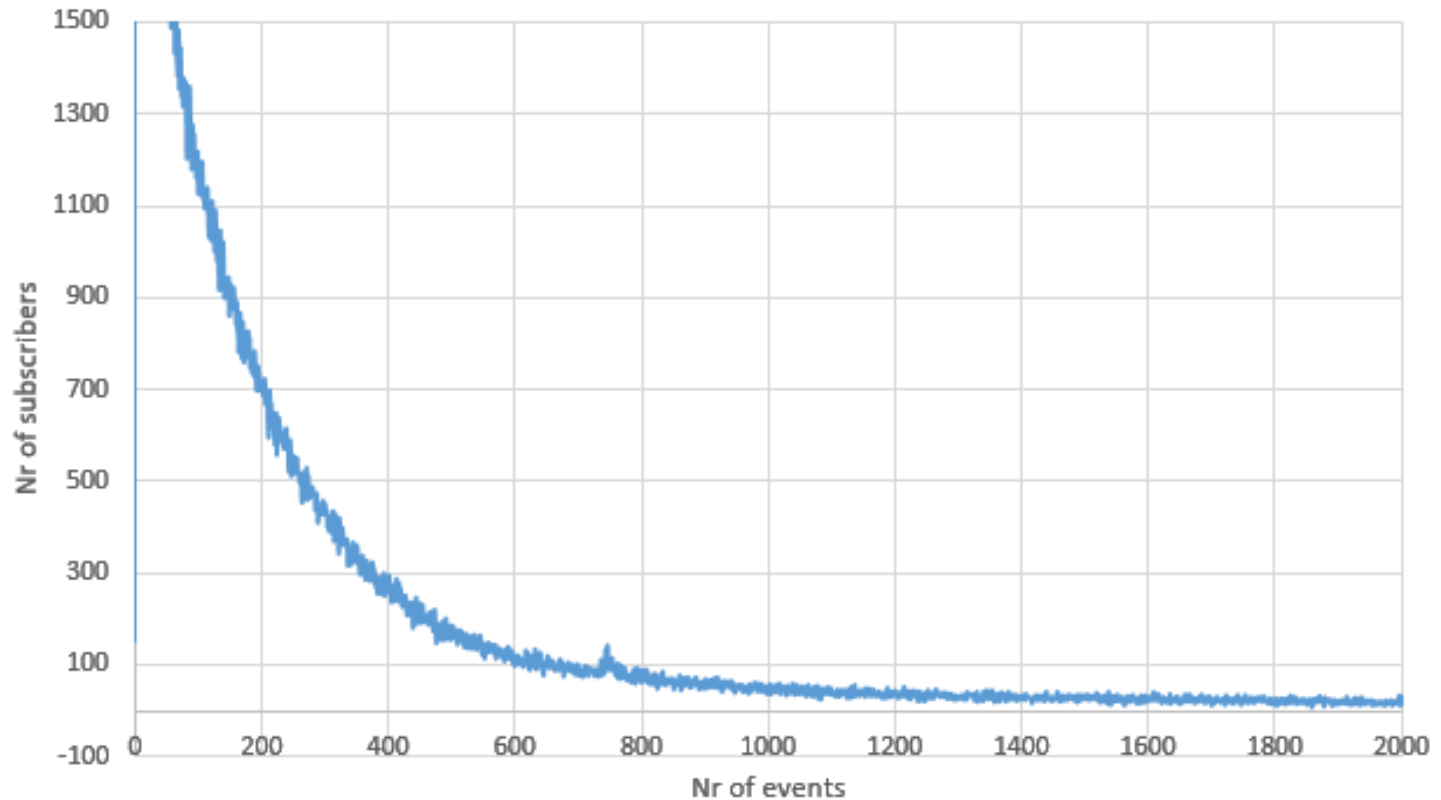
# Stream data distribution

Distribution of stream events



# CDR data distribution

Distribution of CDR events



# Cellplan

- Initial data
  - Coordinates of the cell
  - Start and end angle
  - Technology
    - 2G
    - 3G
    - 4G
  - Radiuses are missing and generated using Voronoi (for 2G)



# Test group

- Events collected for subscribers
  - 12 stream
  - 11 CDR
- Manually collected actual home and work locations of test group

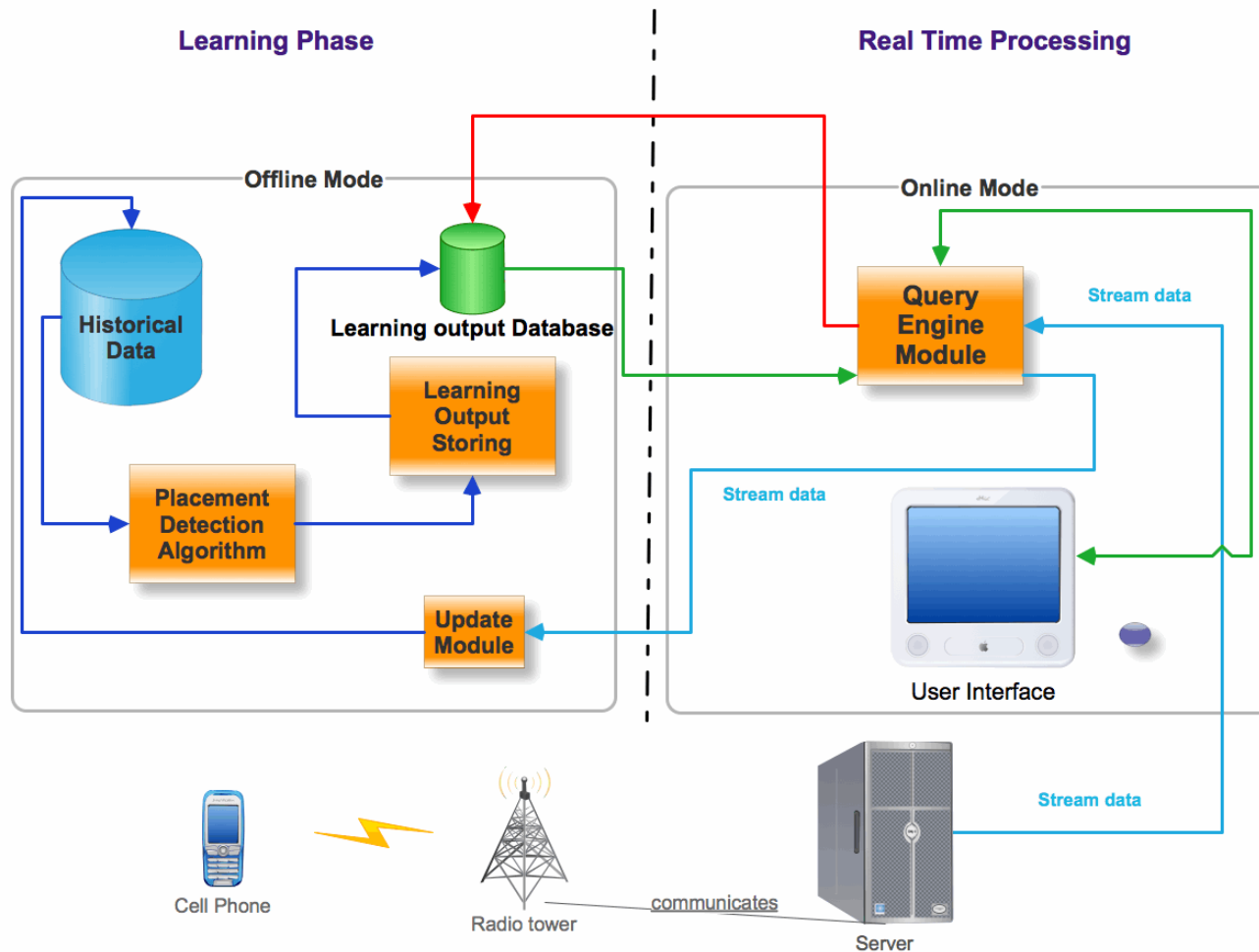


# Location detection

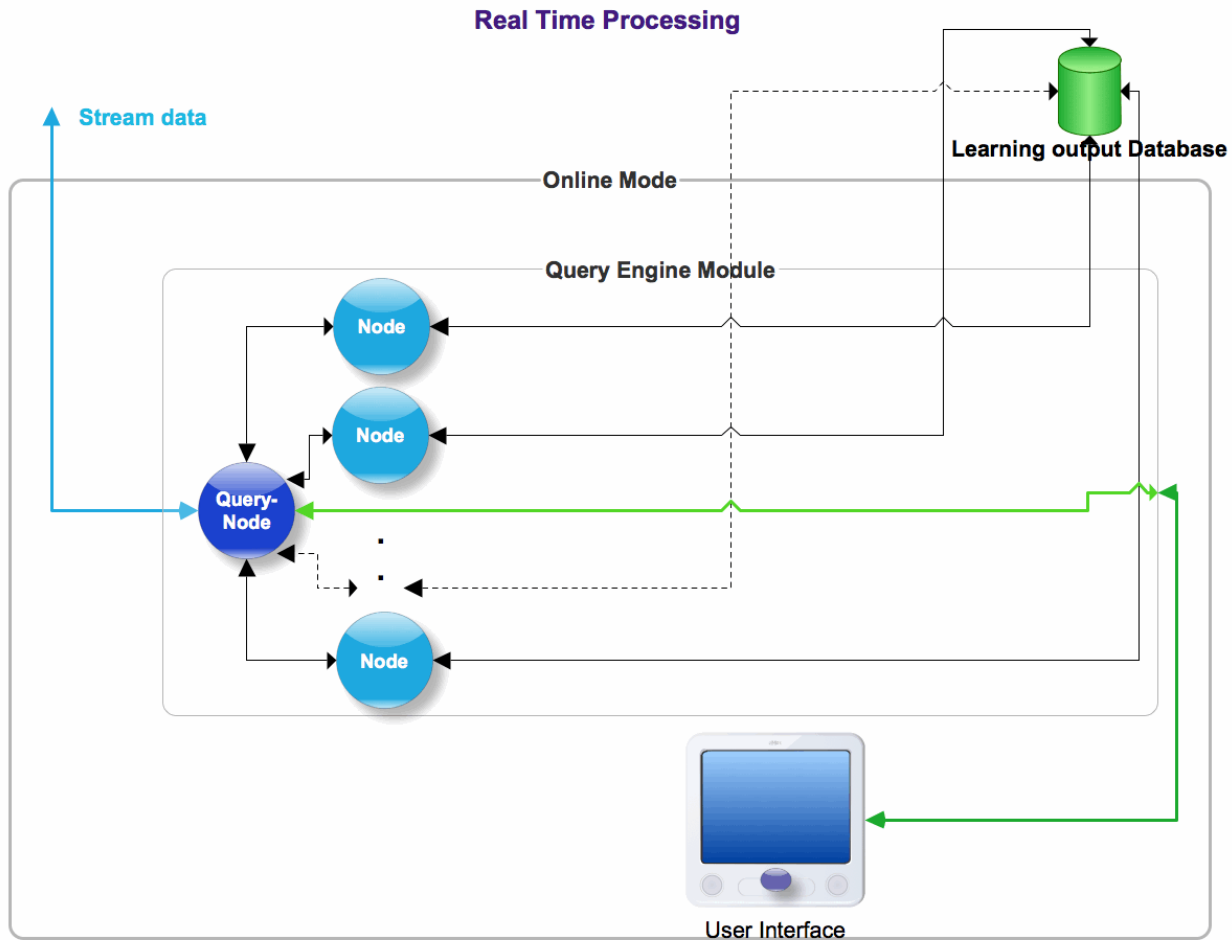
- The idea was inspired from [Ahas 10]
- Gather events for each subscriber from specific hours
  - Home – from 18 until 4
  - Work – from 9 until 16
- Calculate home and work locations every day, using 30 days of data

(\*) Rein Ahas , Siiri Silm , Olle Järv , Erki Saluveer & Margus Tiru (2010) Using Mobile Positioning Data to Model Locations Meaningful to Users of Mobile Phones, Journal of Urban Technology, 17:1, 3-27, DOI: [10.1080/10630731003597306](https://doi.org/10.1080/10630731003597306)

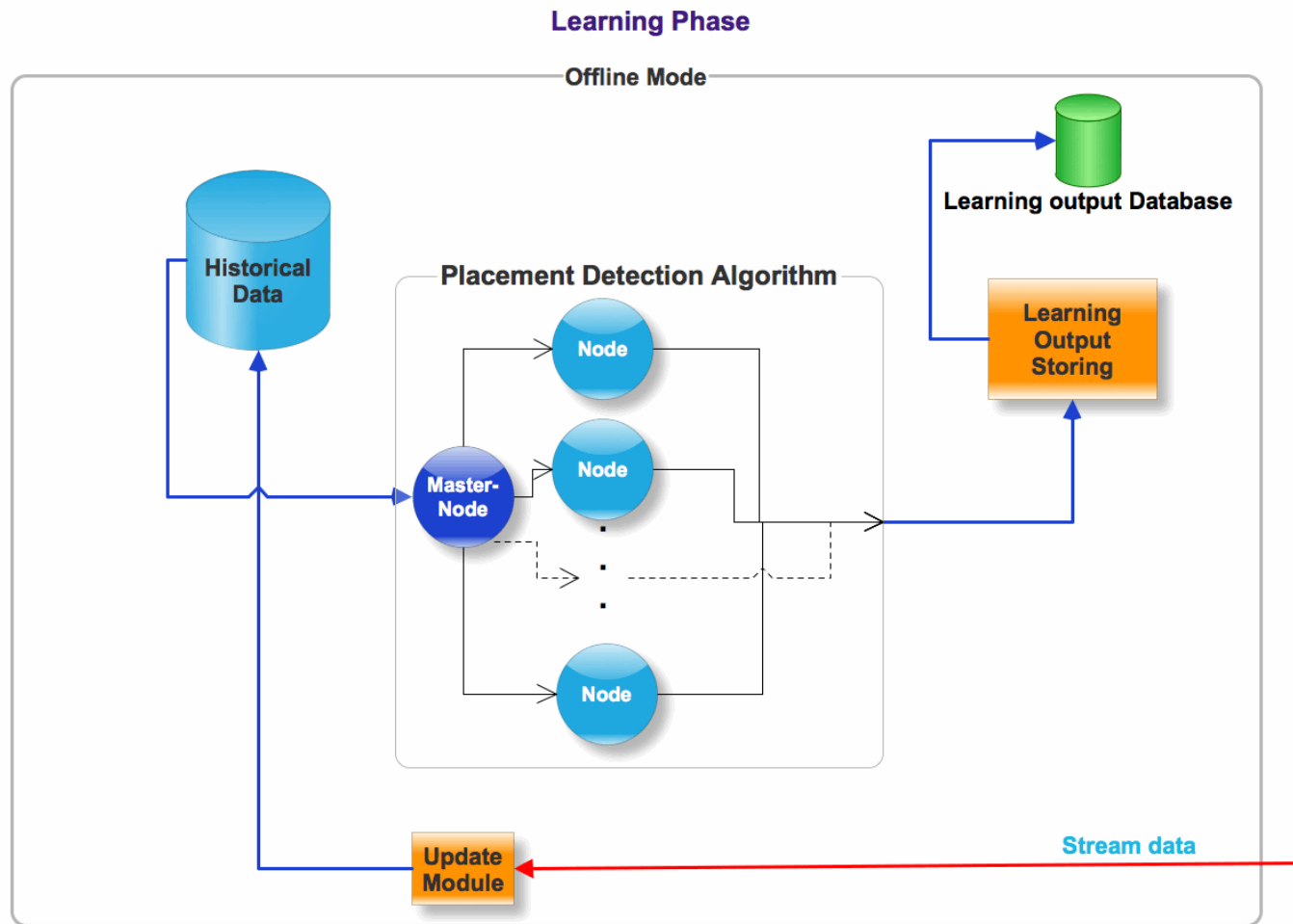
# Real time labeling overview



# Real time labeling – online mode

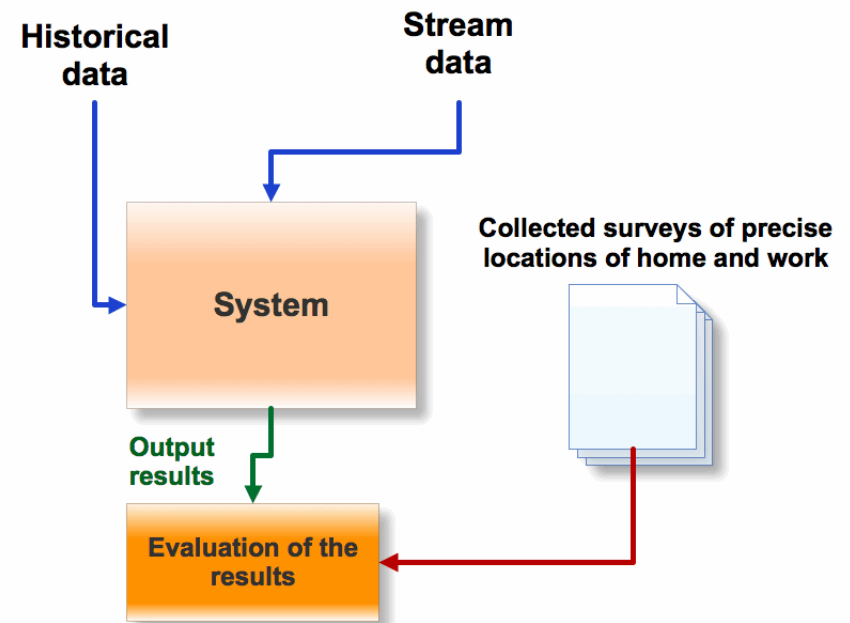


# Real time labeling – offline mode



# Experiment & results

- Accuracy
  - How the accuracy changes when more data is added?
  - 1w, 2w, 4w, 5 (stream)/10 (CDR) months time periods
  - Compute distance difference
- Speed
  - 1 month data, 1M subs, database based solution – ~14 hours
  - 1 month data, 1M subs, JAVA - ~2 hours
  - How the speed improves when distributing data and calculation between nodes



# Test group averages of stream data

Averages of distances and events of our test group

	Distance diff avg (m)	Total events avg	Events in hours avg	CGI count avg
Home_7_days	4848.35	266.5	84	46.08
Home_14_days	659.8	585.83	167.33	82
Home_28_days	639.43	1109.75	339.83	141
Home_5_months	342	6415.33	1910.75	632.58
Work_7_days	691.55	266.5	73.66	40.41
Work_14_days	691.55	585.83	143.5	75
Work_28_days	6508.8	1109.75	275.91	131.83
Work_5_months	573.16	6415.33	1696.83	735.08

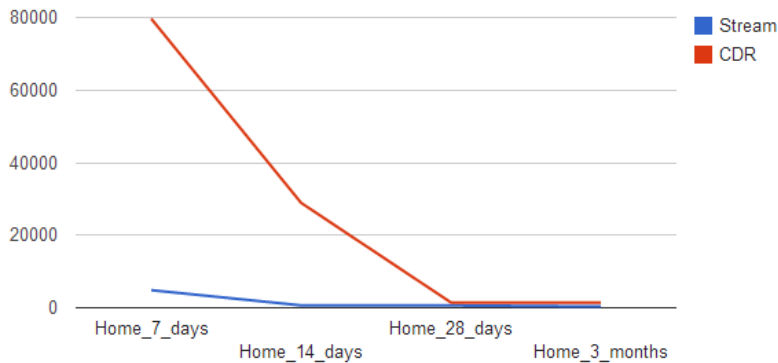
# Test group averages of CDR data

Averages of distances and events of our test group

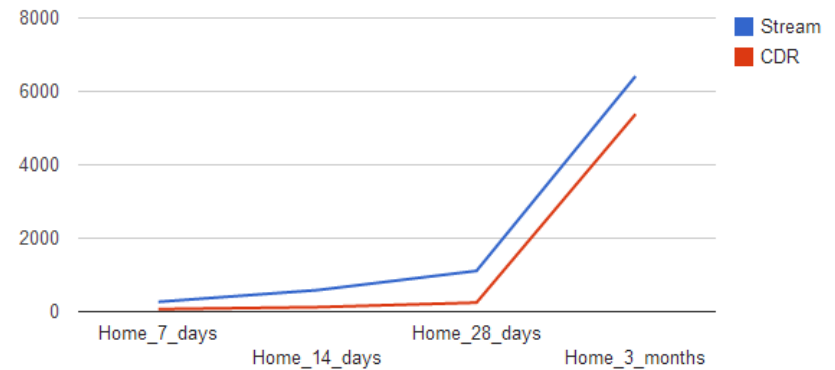
	Distance diff avg (m)	Total events avg	Events in hours avg	CGI count avg
Home_7_days	79673.32	68	18.16	10.66
Home_14_days	28908.9	122.375	30.12	18.5
Home_28_days	1438.2	245	66.62	35.25
Home_10_months	1374.17	5380.62	1496.37	978.62
Work_7_days	31802.86	52.25	21.37	18.12
Work_14_days	1194.02	122.37	51.25	45.37
Work_28_days	844.72	245	95	89.5
Work_10_months	1089.99	5380.62	1579.5	1036

# Results graphs

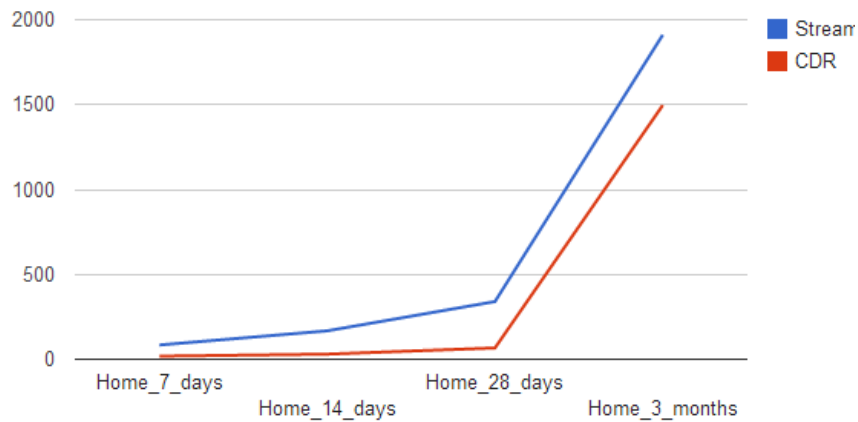
### Distance difference average



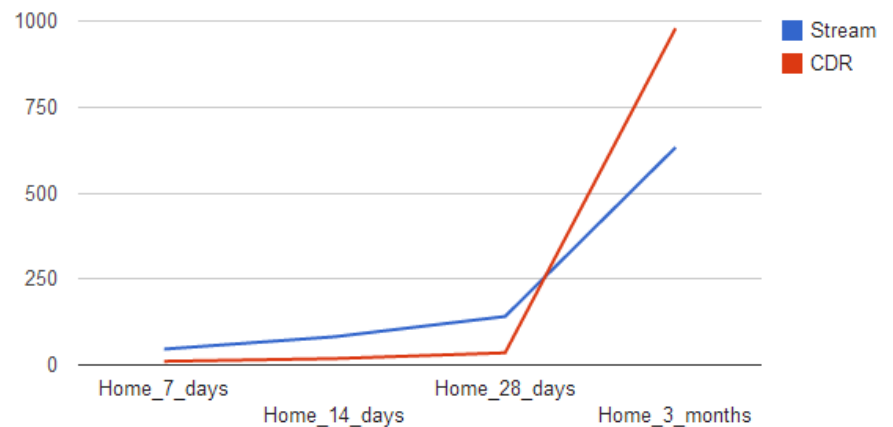
### Total events average



### Events in hours average



### Most frequent CGI count average





# Speed estimation

- 1 million subscribers
  - 1000 events per month for subscriber
  - 4 nodes (1x8Core@2GHz, 64GB RAM)
- Daily home/work calculation (learning)
  - 10 minutes
- Real time labeling takes (real-time)
  - 5-10 ms per event

# Conclusion and Future work

- Our algorithm is suitable for high level home and work detection
- It works with both, stream events and CDR's, though stream provides better results
- Algorithm is scalable and can be used safely with up to 5 millions of subscribers
- Increase the number of test group to 200
- More complex location detection
  - Places where people work out, shop, study
- Subscriber profiling:
  - Who are schoolchildren
  - Who attend sporting events

# Demograft demos

- Targeter:

<https://demo.demograft.com/public/>

- Mobile Broadband Promoter

<https://demo.demograft.com/public/mbp>

- Network Customer Experience

<https://demo.demograft.com/public/nce>

# Thank You! Questions?