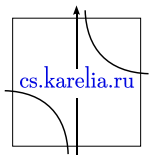
Evaluation of the Smart Space Approach in Mobile Data Processing

Sergey Marchenkov, Pavel Vanag, Dmitry Korzun

Petrozavodsk State University
Department of Computer Science



The project was supported by grant KA179 “Complex development of regional cooperation in the field of open ICT innovations” of Karelia ENPI CBC programme 2007–2013 of the European Union, the Russian Federation and the Republic of Finland

The 11th FRUCT Conference, April 23–27, St.-Petersburg, Russia

Table of Contents

Lausanne Data Collection Campaign by NRC–Lausanne

- Nearly 200 individuals
- Duration of 17 months from Sep. 2009
- Nokia N95 mobile phone were used to collect the data
- Continuously collected spatial data
- Social interaction data
- Phone use data

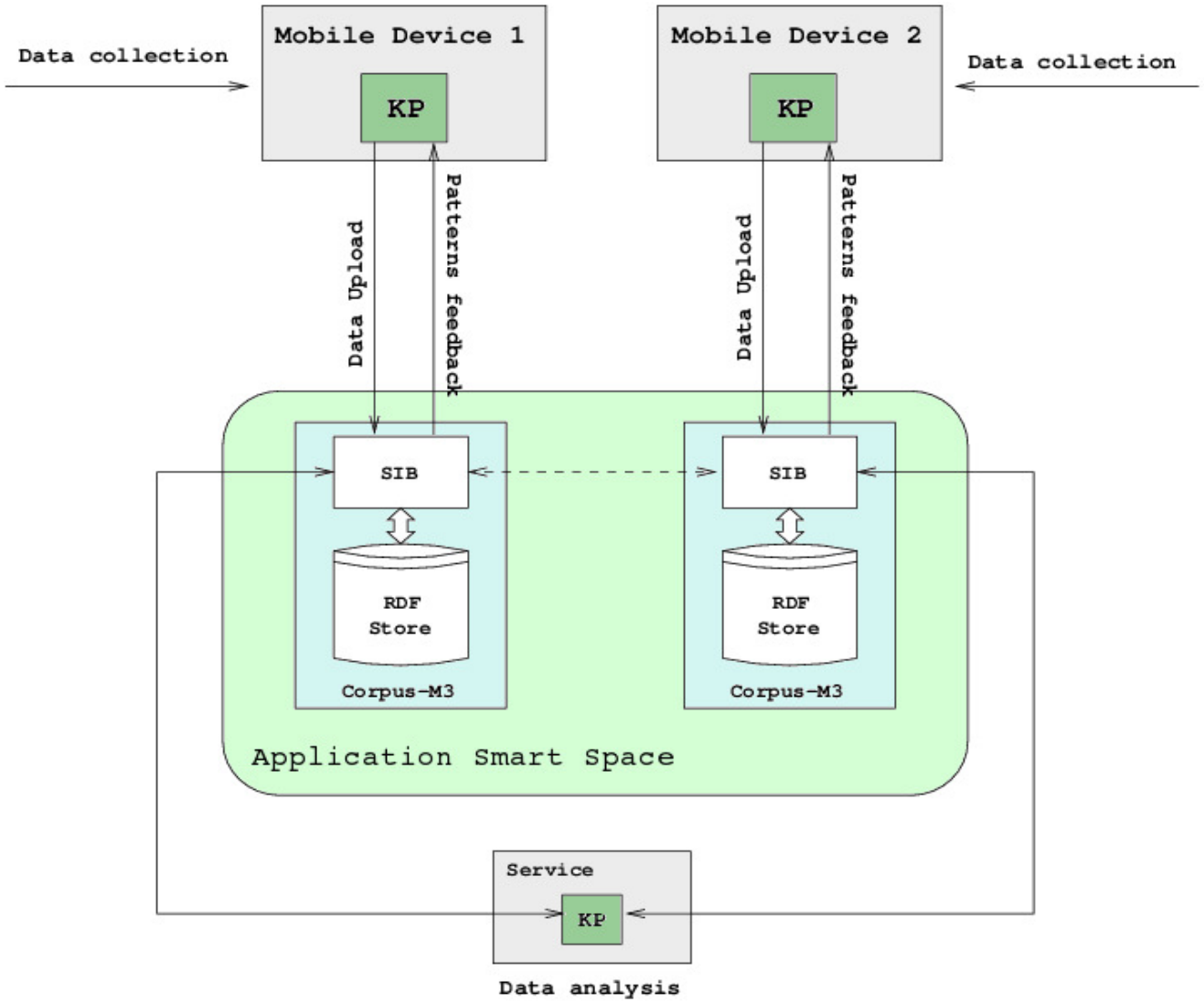
Mobile Data Challenge by NRC–Lausanne

- Releases the Lausanne data (MDC Data Set) made available for the research community in Jan. 2012
- Two alternative tracks: Open Track and Dedicated Track
- Around 700 individual researchers registered, almost 500 challenge tasks were registered
- “Mobile Data Challenge by Nokia” workshop in connection with Pervasive 2012 (June 18th 2012, Newcastle, UK)

Characteristics of the MDC Data Set

- Data on 38 individuals, Sep. 2009 — Apr. 2011
- Data size 17.7 GB
- 13 types of data:
  - accel – scan of the accelerometer sensors
  - application – application events
  - bluetooth – bluetooth devices seen by the user
  - calendar – calendar entries
  - calllog – calls log
  - contacts – contact entries
  - gps – gps positions of the user
  - gsm – gsm cells that the user has seen
  - media – media found on the device
  - mediaplay – information on how user had play media
  - process – informations on the running processes
  - sys – general system informations about the phone
  - wlan – wlan devices seen by the user
  - wlan_loc – geo-position of the wlan access points

Mobile data processing with Smart-M3

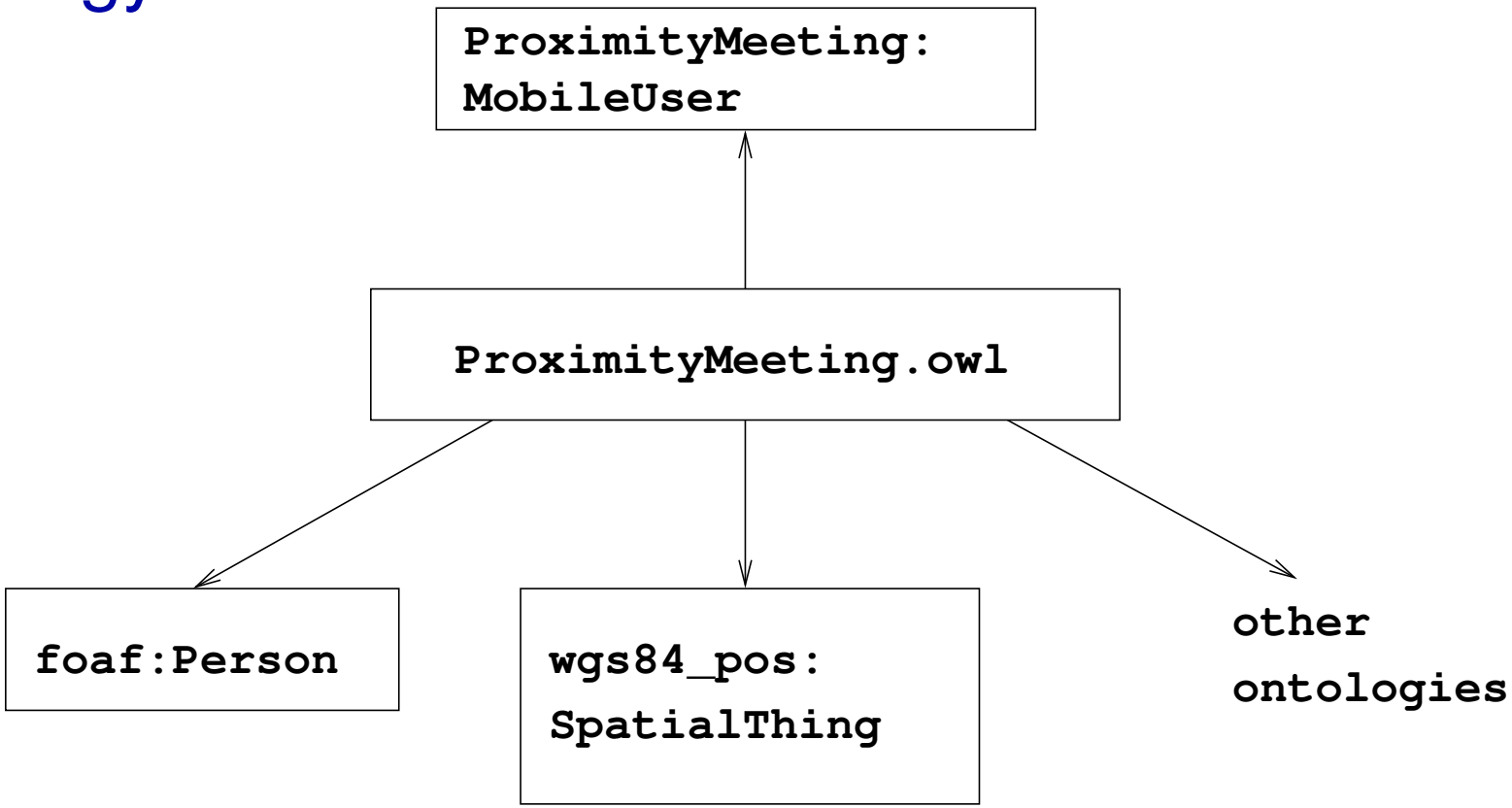


- Distributed computing
- Many mobile participants, a lot of data
- Ontology-driven data processing
- Dynamic data sharing and knowledge reasoning

Open questions

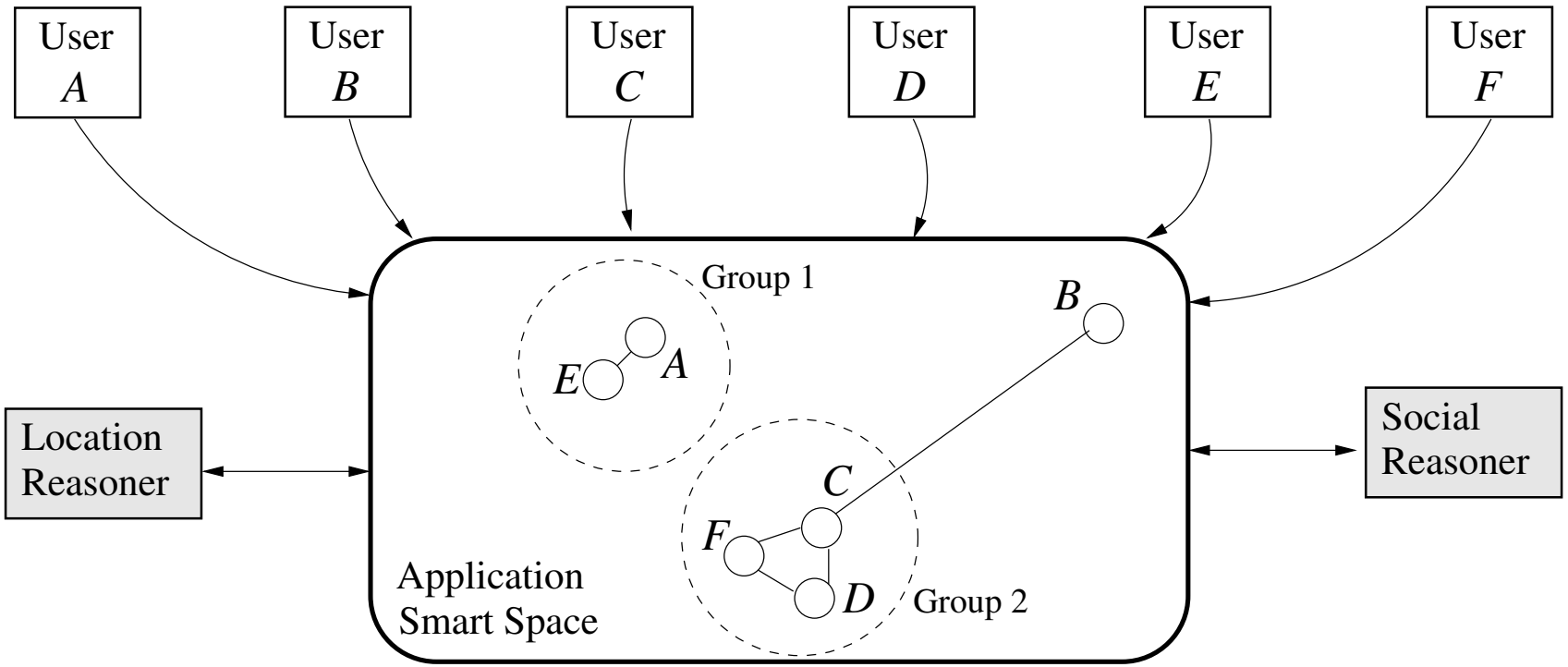
1. Is it possible to apply ontological approach of Smart-M3?
   - behavior traces of people living in their everyday lives
   - RDF/OWL representation models
   - operations with huge multi-person data amounts
2. Is it possible to effectively implement such applications?
   - rapid development in problem-domain terms
   - latest version of FRUCT developed Smart-M3 SDK

Ontology



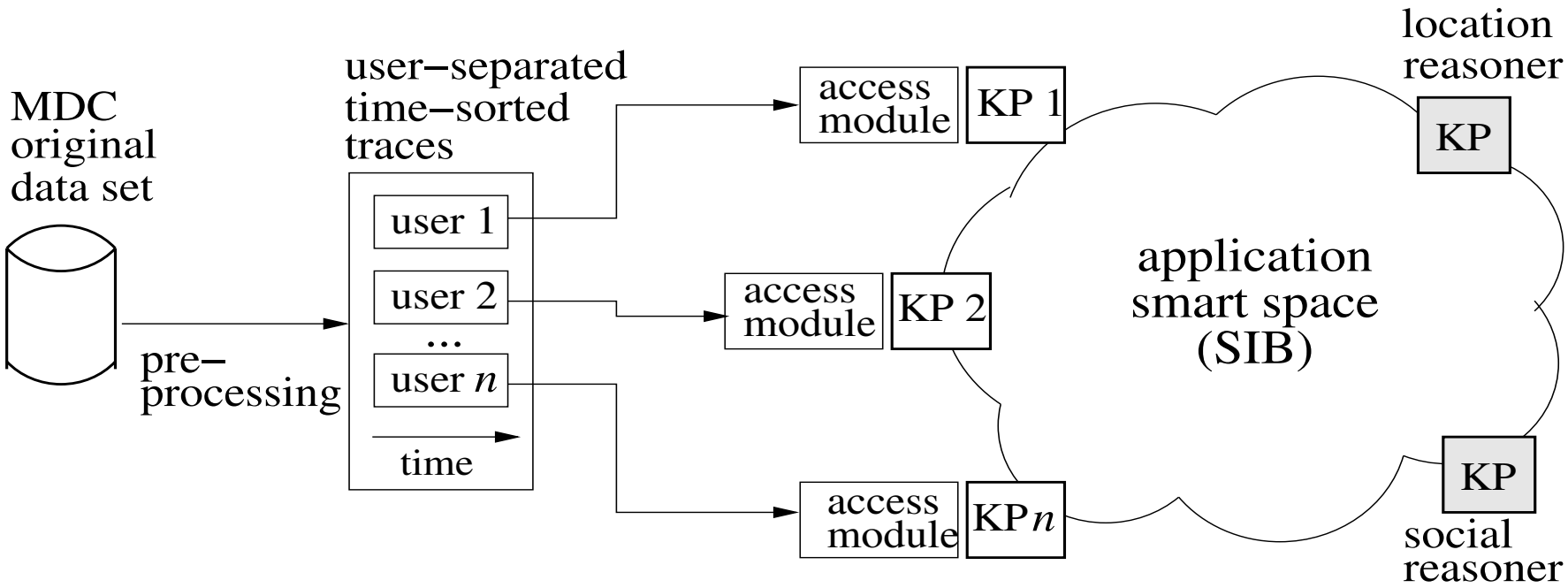
- OWL and standard dictionaries (FOAF, WGS84)
- The modularity allows adaptation to changes
- SmartSlog OWL-to-code generator

Scenario: “Who is near?”



- Mobile users $A, \ldots, F$ share their sensed data
- Reasoners derive knowledge (operationally)

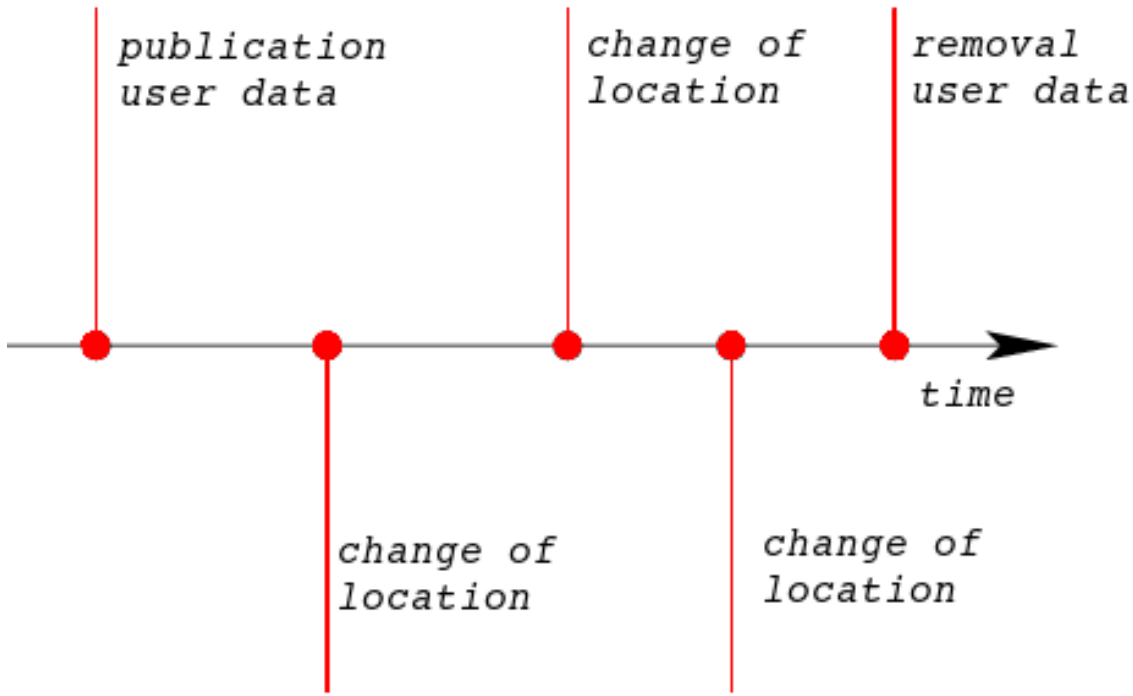
Architecture



- User KP *i* simulates its user by reading own trace produced from the MDC data set for *n* users
- Reasoner KPs operates as in the real scenario

User KP: associated with a mobile user

- Ontology-driven data processing:
  RDF and OWL representation models
- Change over time the smart space: accessing SIB by SSAP



- User KP makes low load: join, insert, updates, leave

Location Reasoner: searching users nearby

- Personalization: reasoning is for a given user
- KP detects when other users are close
  - with a given radius
- Dynamic knowledge reasoning
  - Search queries to the space
  - Subscription
- Focus is on the performance
  - Accessing the data
  - Local reasoning

Samples construction

- A sample is produced for each $n \in [1, 30]$, all run on the same physical machine and start at the same time
- User KP reads the next event data every δ seconds $\delta = 5, 10$ seconds and at least 100 read events
- User KP
  - U_{loc}: local transformation "sensed data $\rightarrow$ OWL/RDF"
  - U_{ss}: access operation for publishing the data to the space
- Reasoner KP
  - R_{loc}: search query construction locally
  - R_{ss}: access operation for resolving the query
  - R_{rsn}: local reasoning based on the reply
- Median ($Q_{50\%}$), average and percentiles ($Q_{10\%}$, $Q_{25\%}$, $Q_{75\%}$, $Q_{90\%}$)

Local processing performance: U_{trs} and R_{ptr}

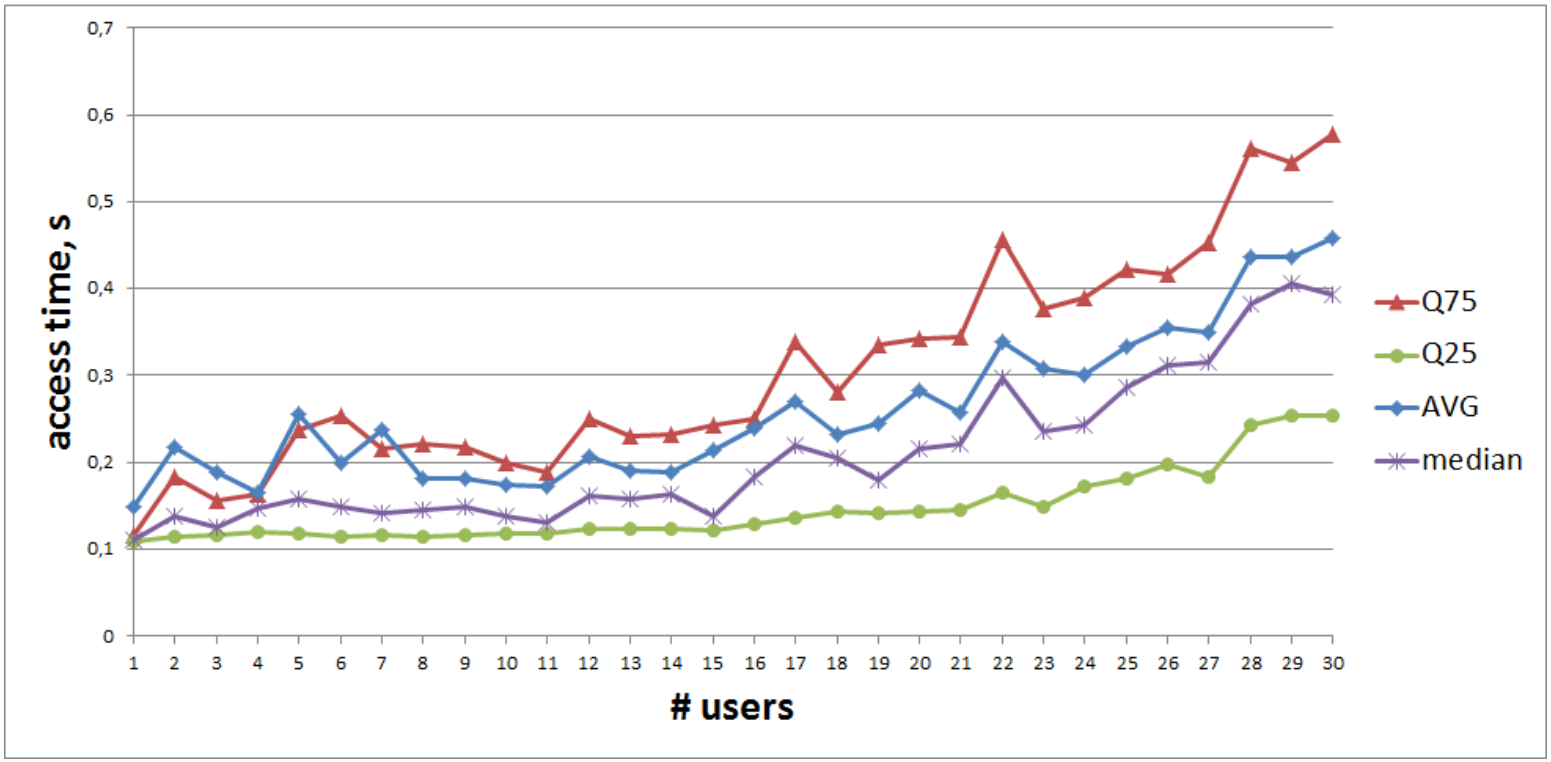
- No significant dependency on the number of participating users n
- Median, average and percentiles for each sample for given n, then averaged over all $n \in [1, 30]$
- Translation of a small set of personal sensed data (U_{trs}) and construction of knowledge pattern (R_{ptr}) are fast and comparable each with another

	$Q_{90\%}$	$Q_{75\%}$	median	average	$Q_{25\%}$	$Q_{10\%}$
U_{trs}	0,7733	0,0264	0,0206	0,1978	0,0184	0,0172
R_{prn}	0,7984	0,0437	0,0357	0,2211	0,0314	0,0286

- Median is lower than average: the majority of samples is fast and there are a few high picks
- User KP can run on low-performance mobile devices like phones

Access time of user KP: U_{ss}

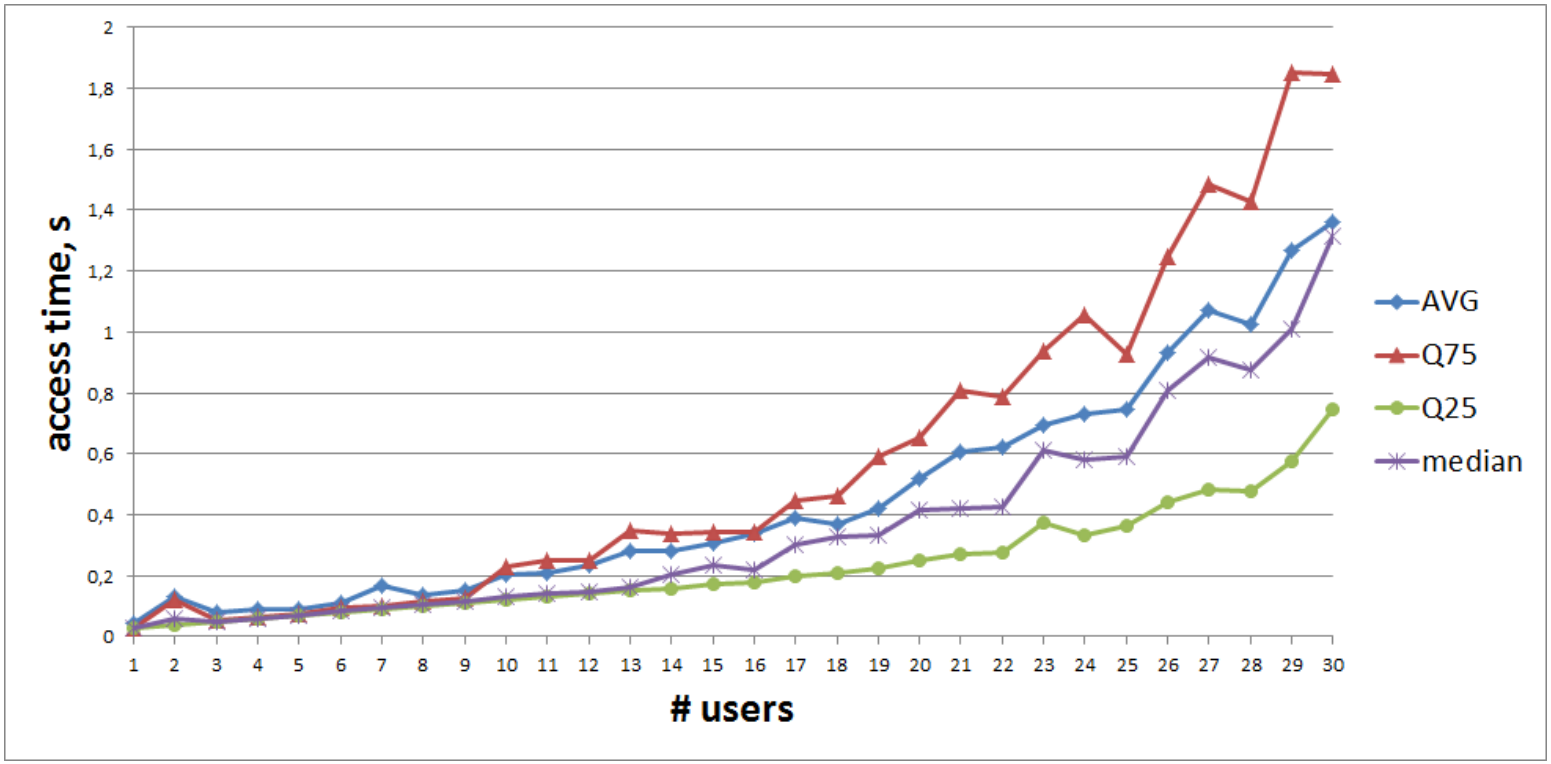
- User KP publishes a portion of personal sensed data



- Dependence on the number of users n, slow growth
- Relatively small, less than 1 second

Access time of reasoner KP: R_{ss}

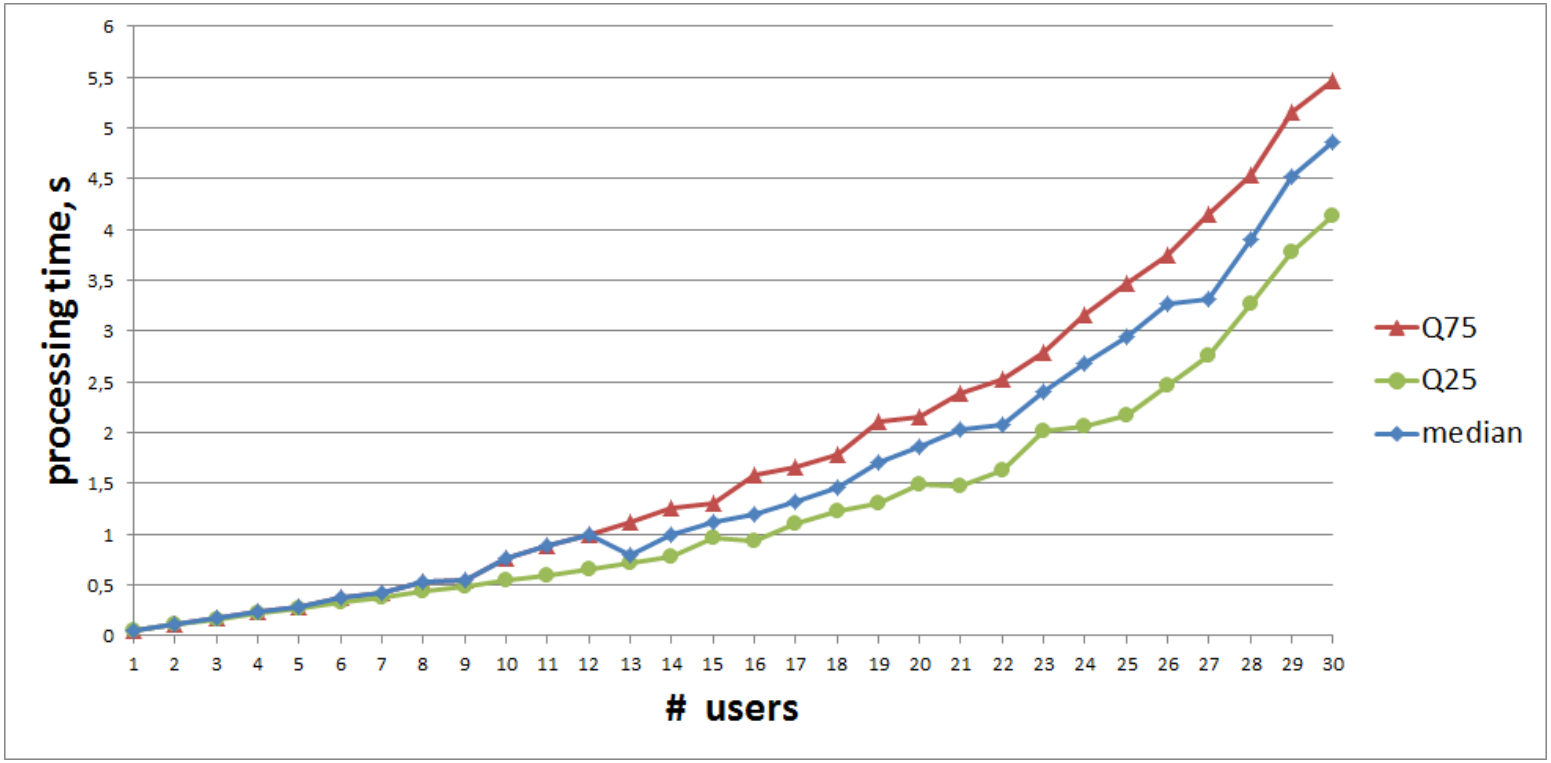
- User KP queries the space for locations of all active users



- Dependence on the number of users n, growth
- About 1…2 seconds

Local reasoning time: R_{rsn}

- Reasoner KP analyzes queried data to find the users nearby



- Dependence on the number of users n, essential growth as $O(n^{1+\epsilon})$ for small $0 < \epsilon < 1$
- The worst-case scenario: reasoning is on the KP side

Conclusion

- The smart-M3 platform is in a phase of research prototyping
- It can be used for applications:
  people centric sensing and many mobile users
- Mobile phones as simple sensing devices
- Mechanisms for processing balance between SIB and KP are needed for more efficient reasoning

Q&A