TOWARDS AMBIENT INTELLIGENT APPLICATIONS USING MODELS@RUN.TIME AND MACHINE LEARNING FOR CONTEXT-AWARENESS

PhD defence of Assaad Moawad

11 January 2016

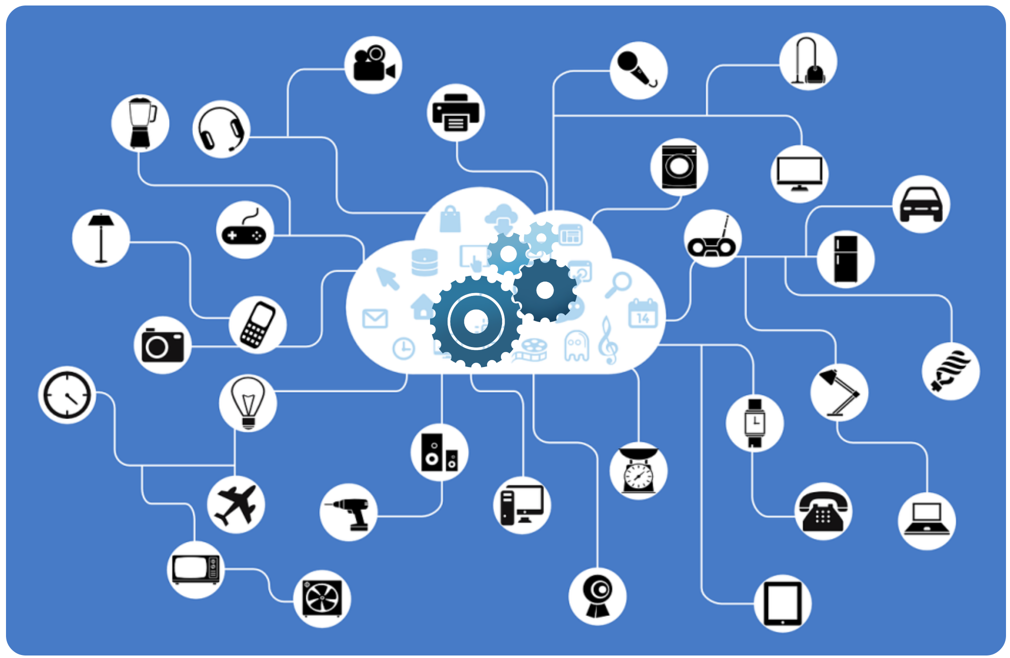
Chairman: Prof. Dr. Nicolas Navet, University of Luxembourg
Co-Chairman: Dr. François Fouquet, University of Luxembourg
Supervisor: Prof. Dr. Yves Le traon, University of Luxembourg
Member: Prof. Dr. Houari Sahraoui, University of Montreal, Canada
Member: A.Prof. Dr. Romain Rouvoy, University of Lille, France
Expert: Dr. Patrice Caire, University of Luxembourg

UNIVERSITÉ DU LUXEMBOURG

SnT securityandtrust.lu

Serval securityandtrust.lu

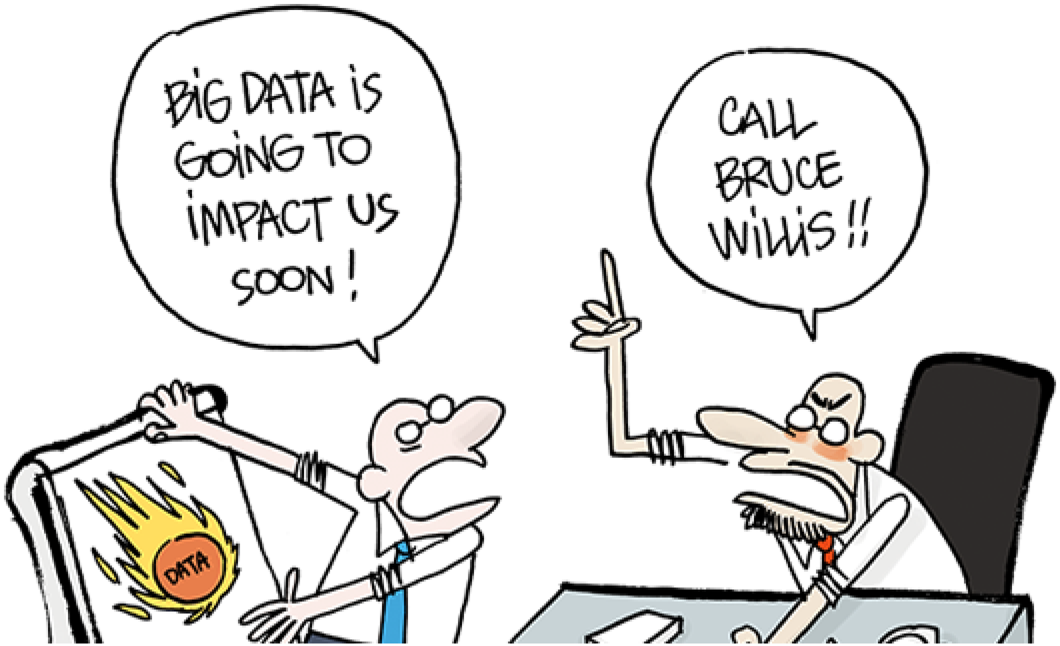
Knowledge comes from data aggregation…



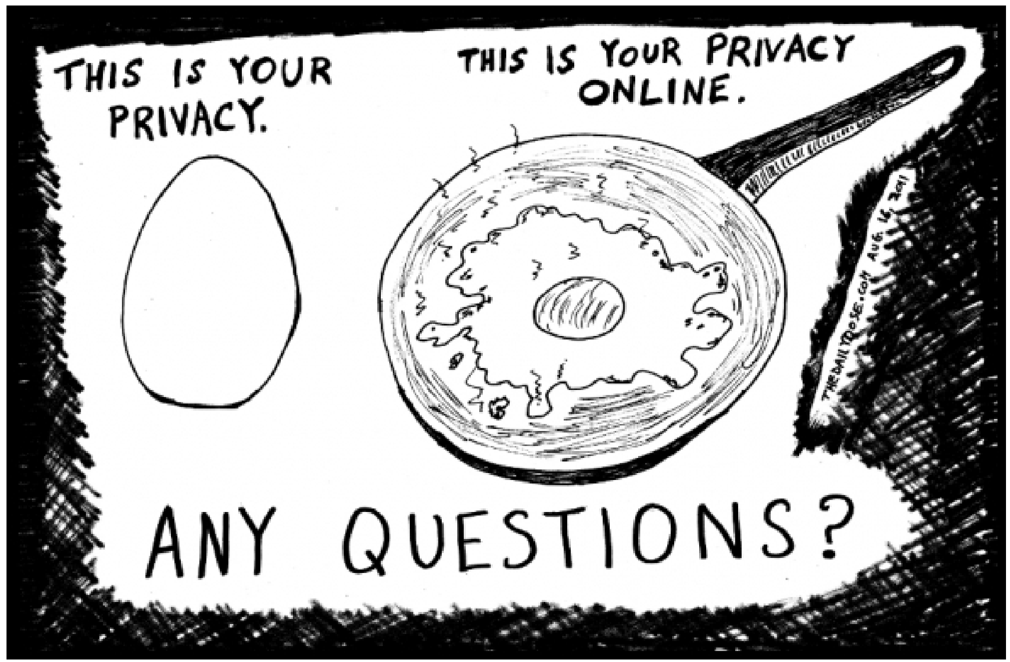
… In order to reason in a particular domain

But... Too much data can dilute knowledge

...this is the big data trap

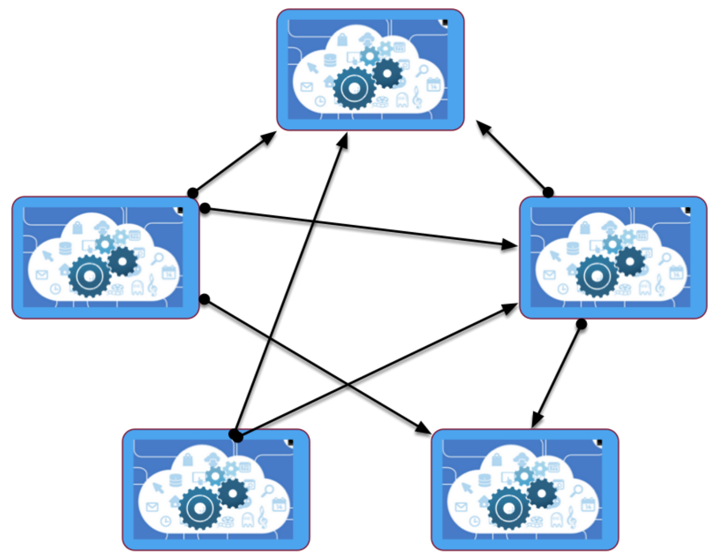


The more data you send...



The higher the privacy risk is

How to avoid the trap?



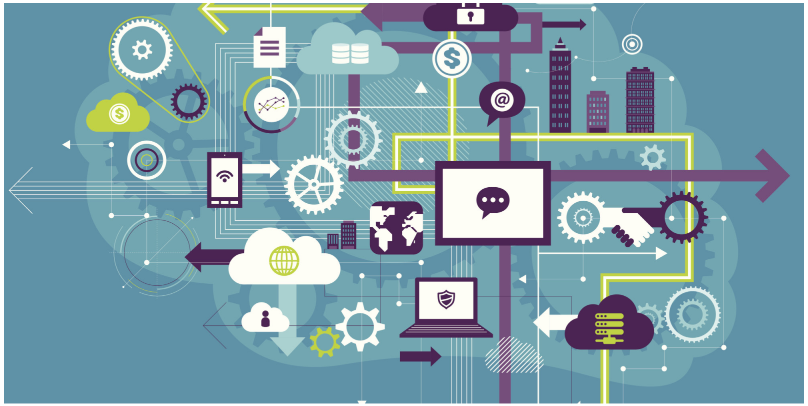
Distributed knowledge models that perfectly fit in reasoning boxes

Part 1 - Application domains

Internet of Things (IoT)

> *Networked* ***interconnection*** *of everyday objects, which are often equipped with ubiquitous intelligence [Atzori et al. 2010]*

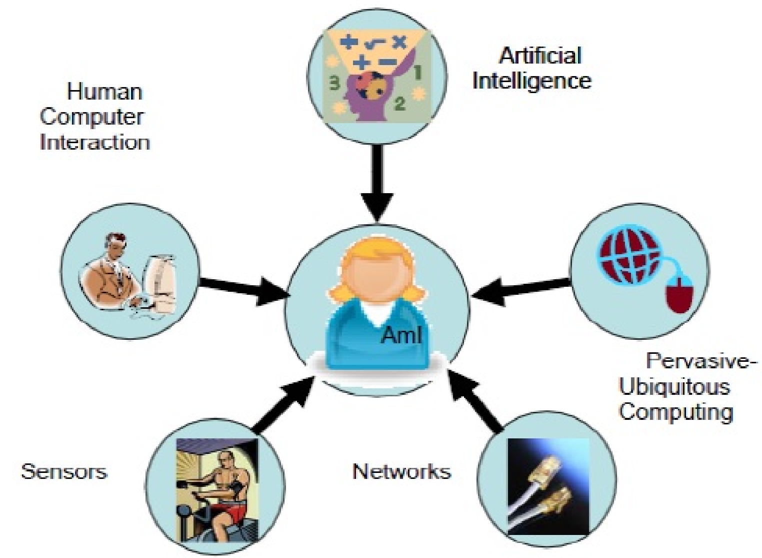
- Ubiquitous communication
- Pervasive computing
- ***Distributed, dynamic and heterogeneous***
- Typically composed of ***smart objects***



Ambient Intelligence (AmI)

Foster a human-machine interaction, where technologies are deployed to make computers disappear in the background [Remagnino et al. 2005]

- **Invisible** interaction with humans
- ***User-centric, adaptive, unobtrusive***
- IoT is one way to realize AmI

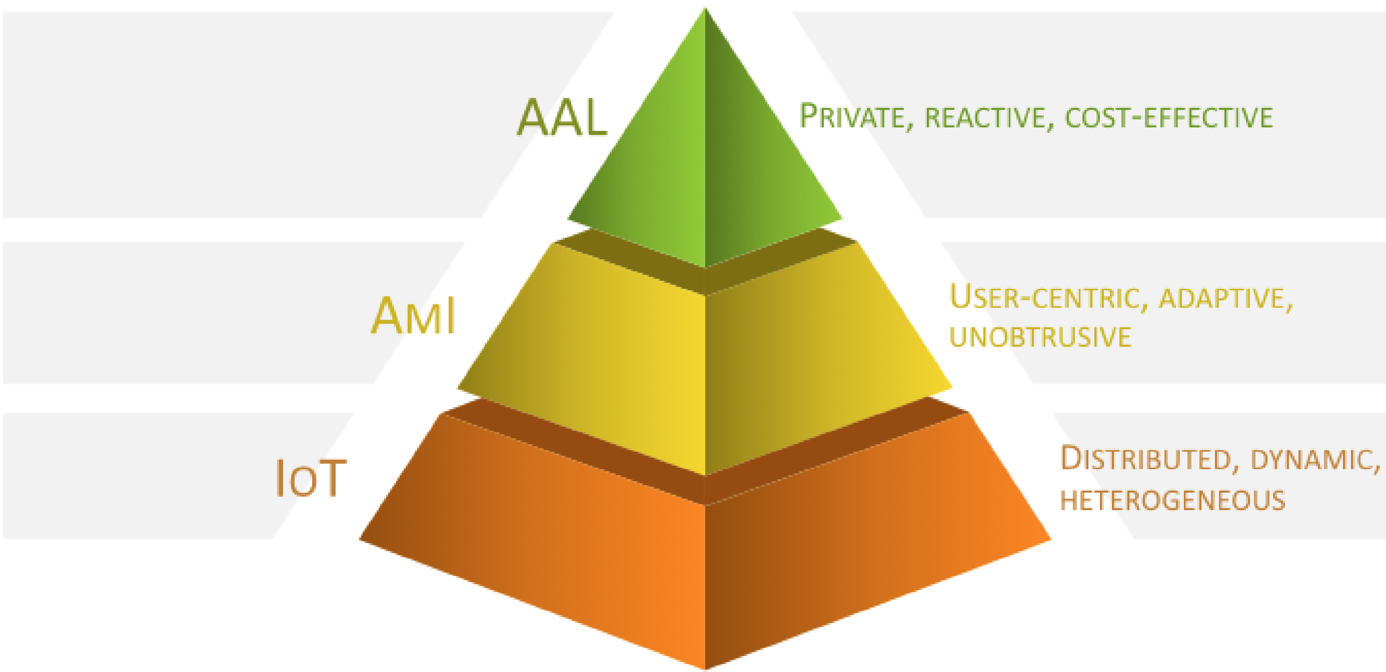


Ambient Assisted Living (AAL)

> *Technical systems to support **elderly people** in their daily routine [Dohr et al. 2010]*

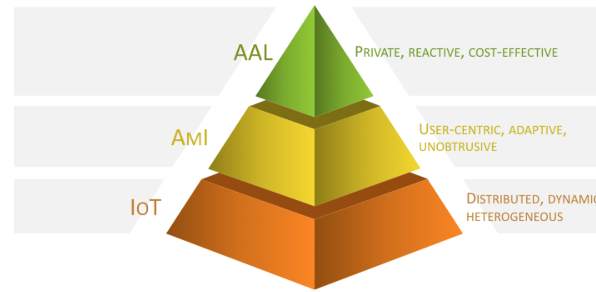
- Critical (**health care** domain)
- ***Private, reactive, cost-effective***
- AAL is a specific case of AmI

Application domains



Background - frameworks

- IoT: *Distributed, dynamic and heterogeneous*



Models@run.time

Kevoree Modeling Framework

Free the code from models !

Component-based middleware

Research questions

- ***R1:*** How to **efficiently** model physical measurements?

Research questions

- ***R1:*** How to **efficiently** model physical measurements?
- ***R2:*** How to enable **distributed context awareness**?

Research questions

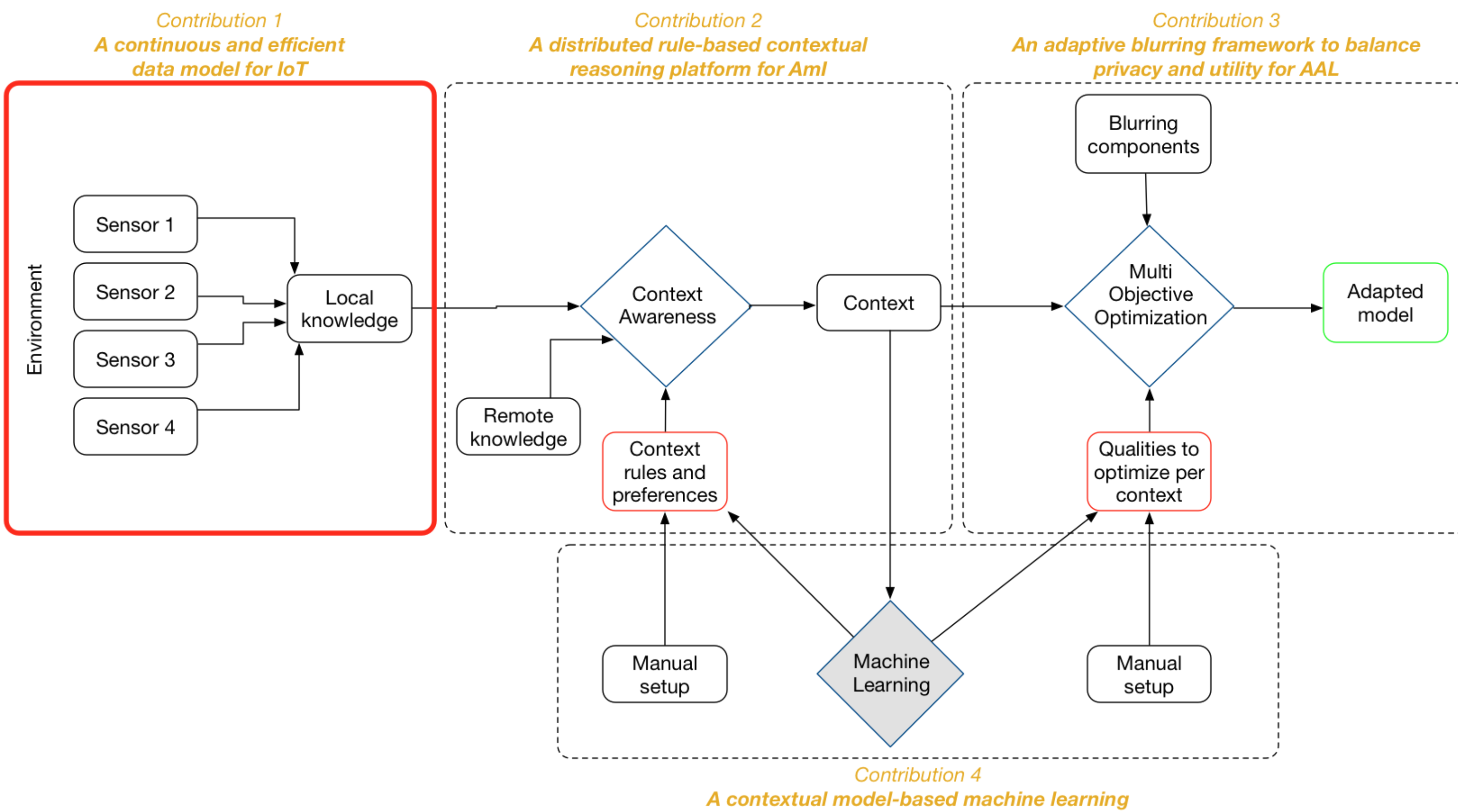
- ***R1:*** How to **efficiently** model physical measurements?
- ***R2:*** How to enable **distributed context awareness**?
- ***R3:*** How to **adapt privacy** when context changes?

Research questions

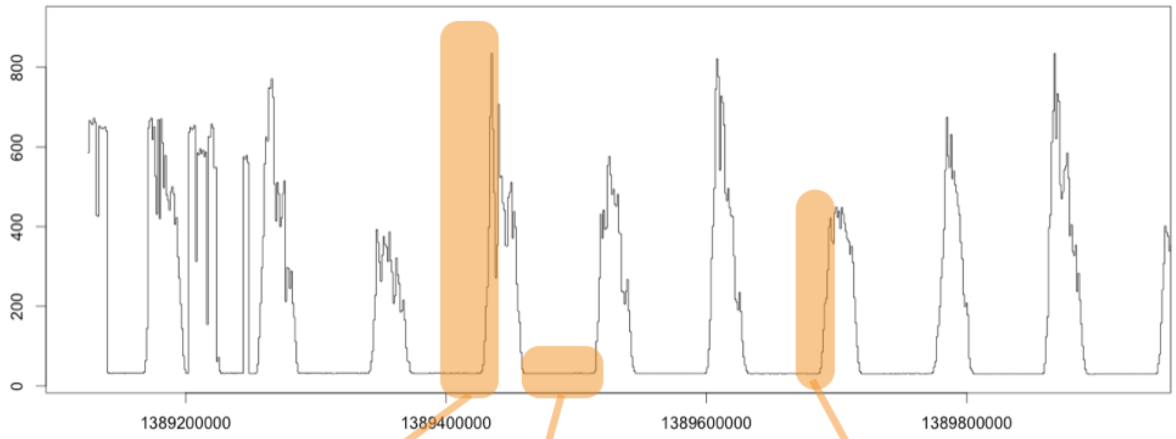
- ***R1:*** How to **efficiently** model physical measurements?
- ***R2:*** How to enable **distributed context awareness**?
- ***R3:*** How to **adapt privacy** when context changes?
- ***R4:*** How to **improve reasoning** using contextual information ?

Part 2 - Contributions

Contribution 1



Reasoning on physical measurements



TimeStamp	Value
07/01/16 16:46	17.63525759
07/01/16 16:47	17.63525759
07/01/16 16:48	17.75439822
07/01/16 17:55	10.52543698

- Intuition: Encode signal as **sequence of polynomials** instead of discrete timestamped values

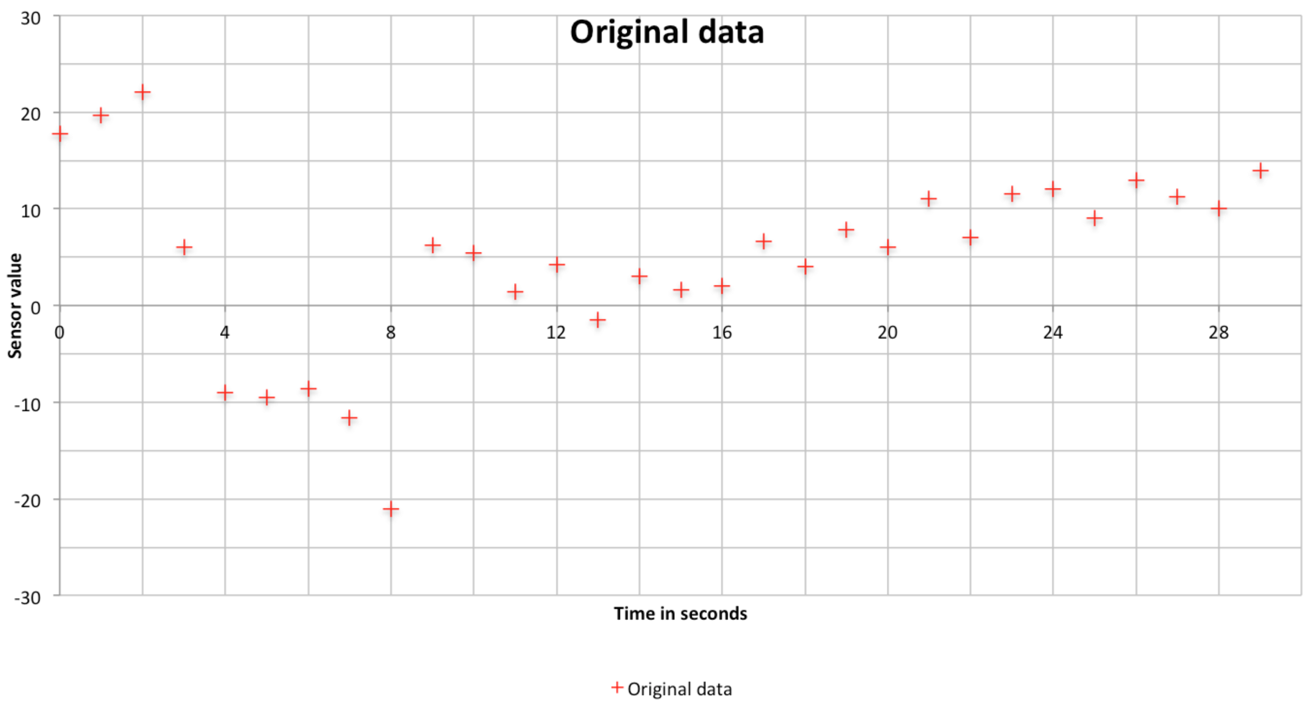
Problem statement

- Physical properties are **continuous** in *time*
- Sampling rate can **vary** (energy saving/network loss)
- Measurements are **imprecise** (sensor precision)

TimeStamp	Value
07/01/16 16:46	17.63525759
07/01/16 16:47	17.63525759
07/01/16 16:48	17.75439822
07/01/16 17:55	10.52543698

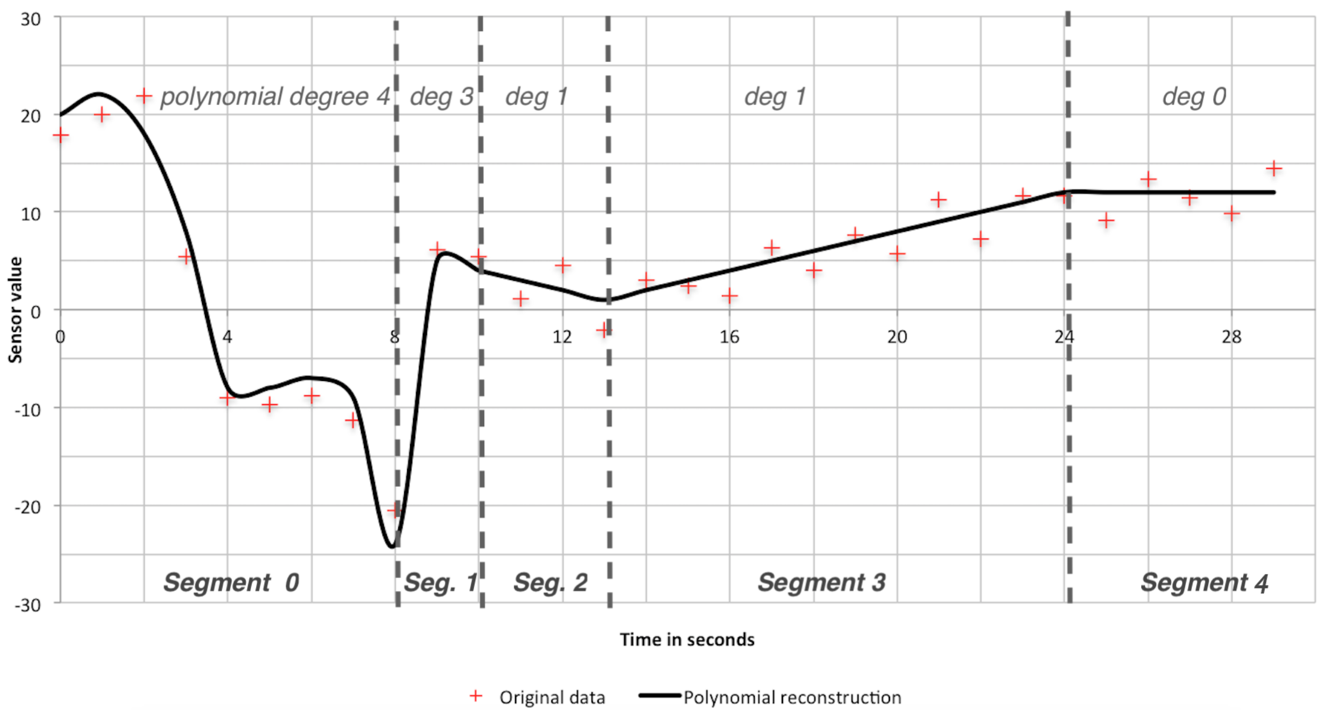
> *“**Models** are supposed to be **cheaper and simpler than the reality**…can we enhance IoT data manipulation by **considering these characteristics** and ultimately speedup **reasoning** (and other) activities on top?*

Example



- Initially 30 points

Live segmentation



- Initially 30 points

  -> 5 polynomials

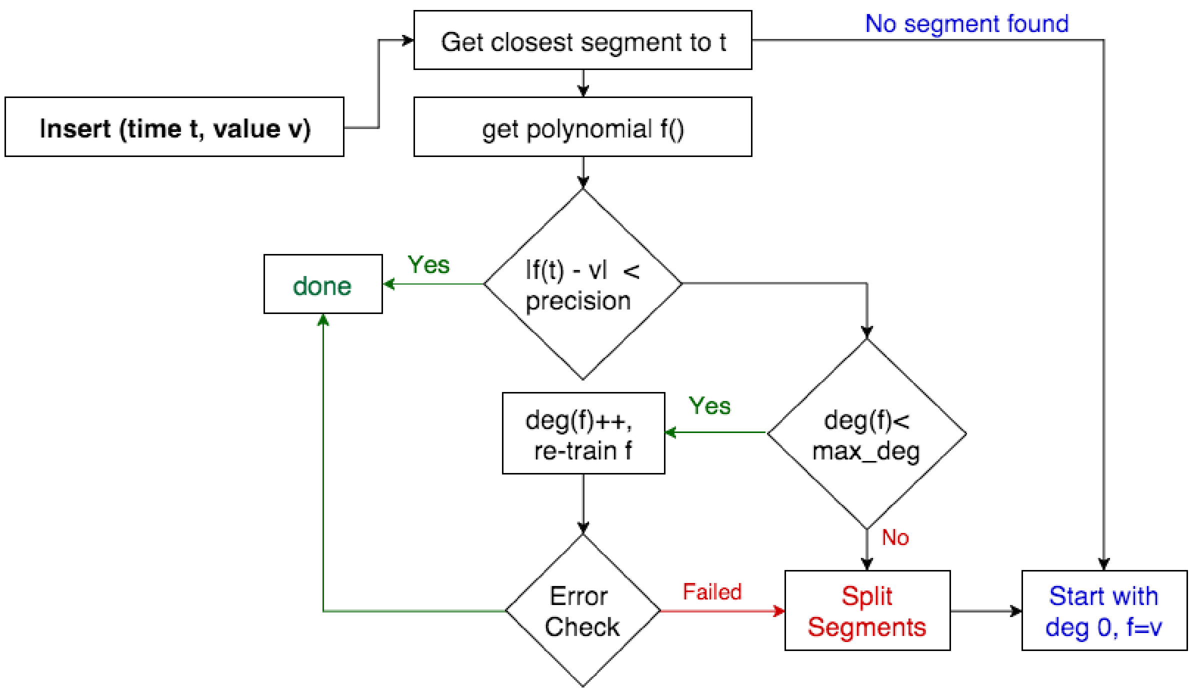
  -> 5 records to store
- 14 doubles to store

  instead of 30

  (14=5+4+2+2+1)
- **Pre-processed data**

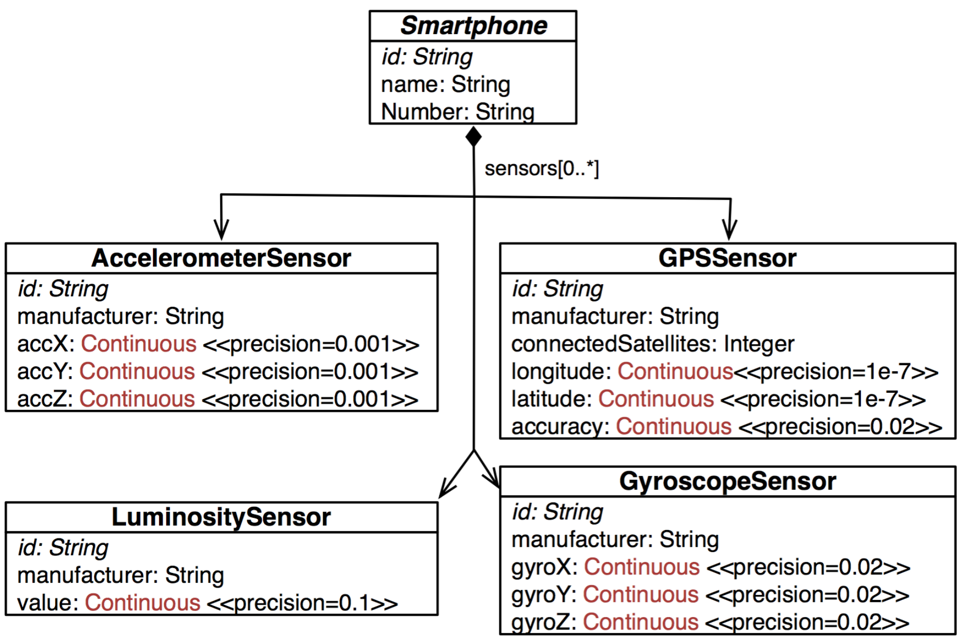
How it works?

> *Insert operation using* ***live machine learning process*** *to build polynomials*



Integration into modeling tools

Continuous meta attributes




```
class smartgrid.SmartMeter extends smartgrid.Entity, smartgrid.Meter {
    att maxAllowedPower: Long
    att electricityActive: Bool
    att highPowerCurrentActive: Bool
    att distance2concentrator: Int
    att hops2concentrator: Int
    att duration2Read: Double

    att activeEnergyConsumedPolynomial: Continuous with precision 0.5
    att activeEnergyProducedPolynomial: Continuous with precision 0.5
    att reactiveEnergyConsumedPolynomial: Continuous with precision 0.5
    att reactiveEneryProducedPolynomial: Continuous with precision 0.5

    func register
    func searchConcentrator : smartgrid.Concentrator
}
```

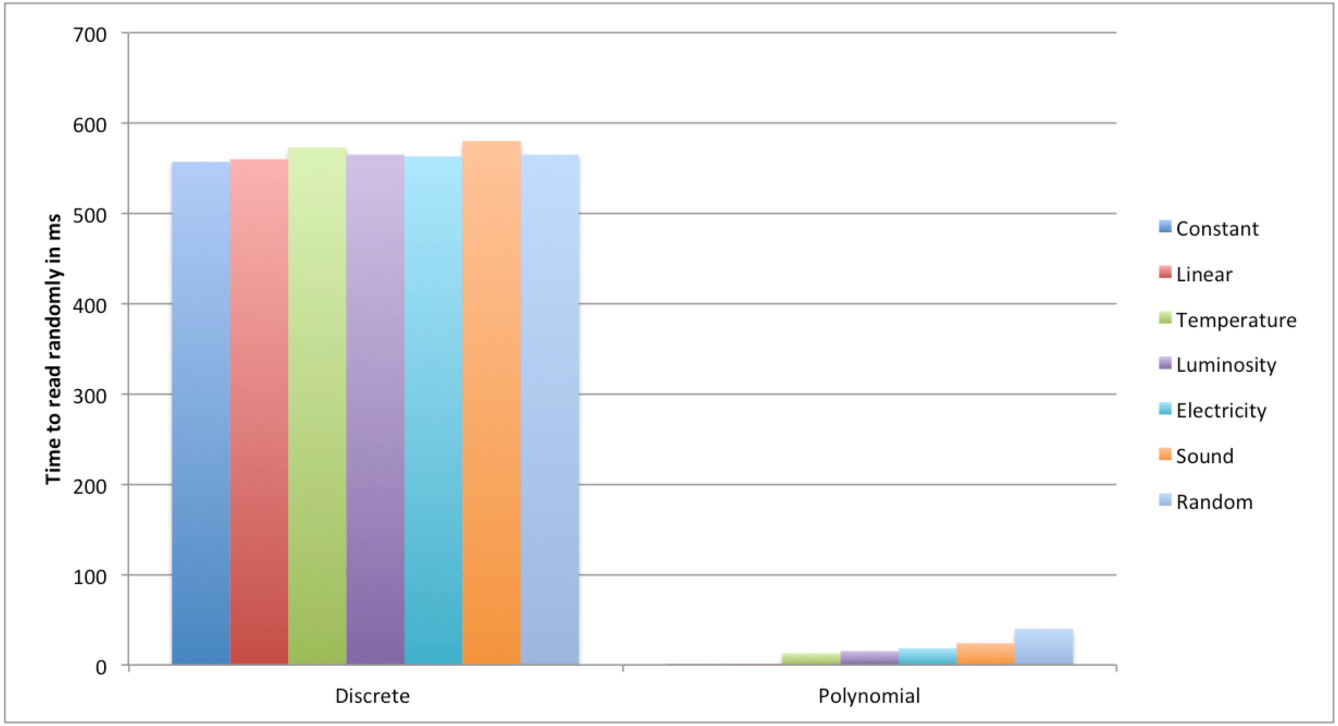
Common experimental protocol

- We define 7 datasets, from the more **constant** to the more **chaotic**
- Each dataset contains **5 000 000** values
- Using KMF 4, **Java** version *(core i7, 16GB, SSD)*, saving to **leveldb**

Database	Sensor
DS1: Constant	c=42
DS2: Linear function	y=5x
DS3: Temperature	DHT11 (0 50'C +/- 2'C)
DS4: Luminosity	SEN-09088 (10 lux prcision)
DS5: Electricity load	from Creos SmartMeters data
DS6: Music file	2 minutes samples from wav file
DS7: Pure random	from random.org

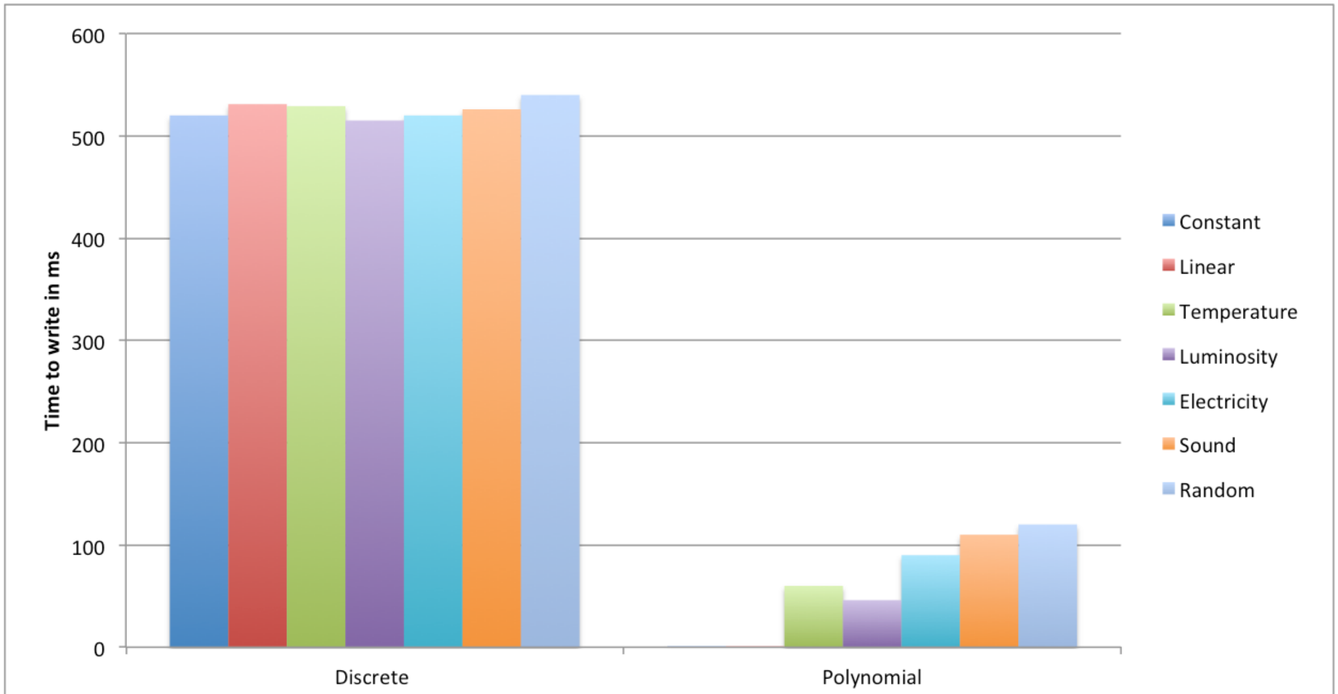
Read operations speed

- **Polynomials** are at least **20-50x faster** than discrete approach



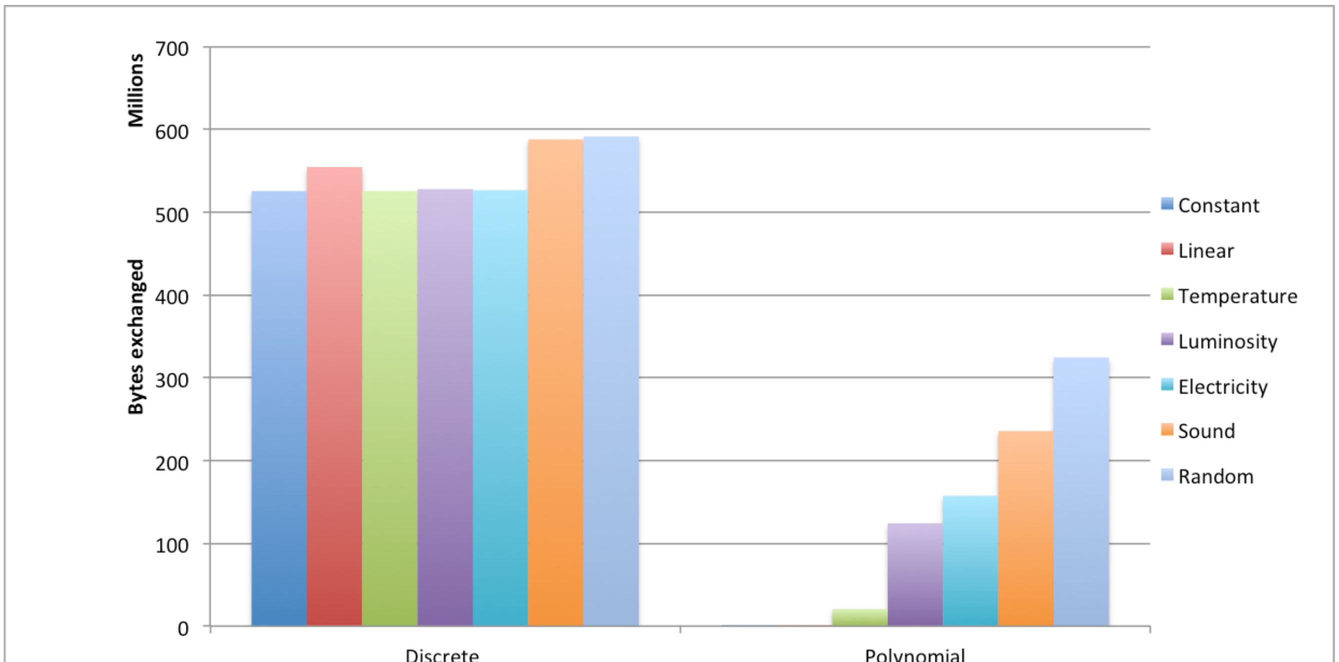
Write operations speed

- **Polynomials** are at least **5 times faster** than discrete approach

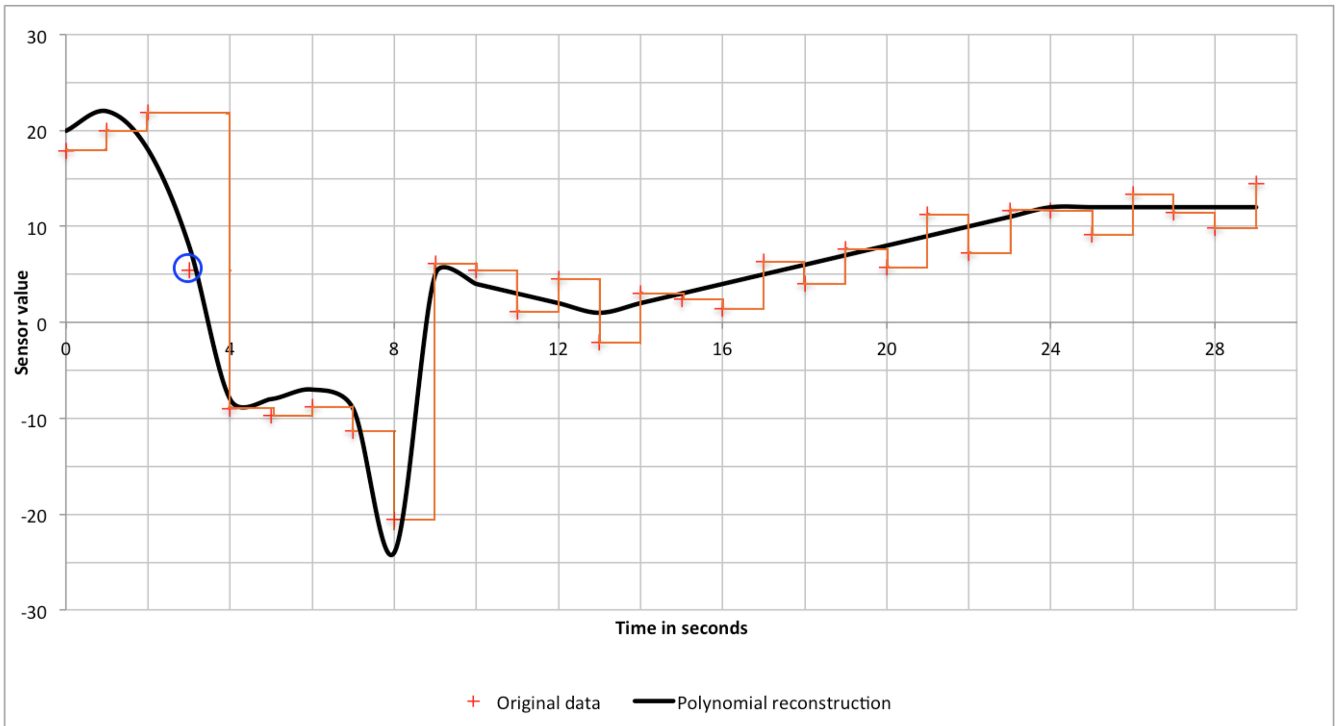


Bytes exchange rate

- Compression rate between **46 to 73%**



Resilience to data loss



Resilience to data loss

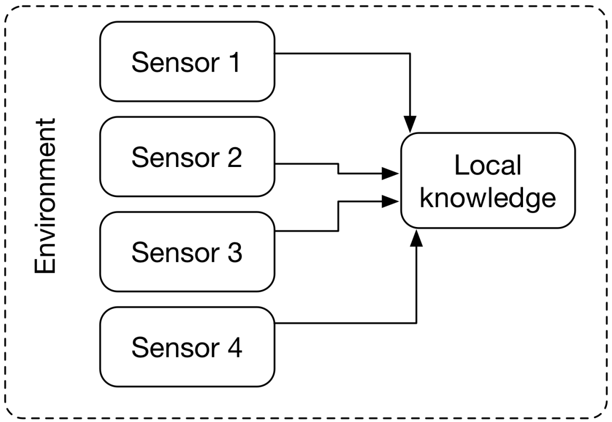
- We **randomly drop 10%** of values in all experimental datasets
- We get less **average error**

Database	Discrete	Polynomial
DS1: Constant	0%	0%
DS2: Linear function	5 %	0%
DS3: Temperature	8.5%	3%
DS4: Luminosity	9.9%	3.5%
DS5: Electricity	17 %	6%
DS6: Sound sensor	21%	13%
DS7: Random	31.8%	30.8%

AVERAGE ERROR WHEN WE TRY TO APPROXIMATE MISSING VALUES

Summary

Contribution 1
A continuous and efficient data model for IoT

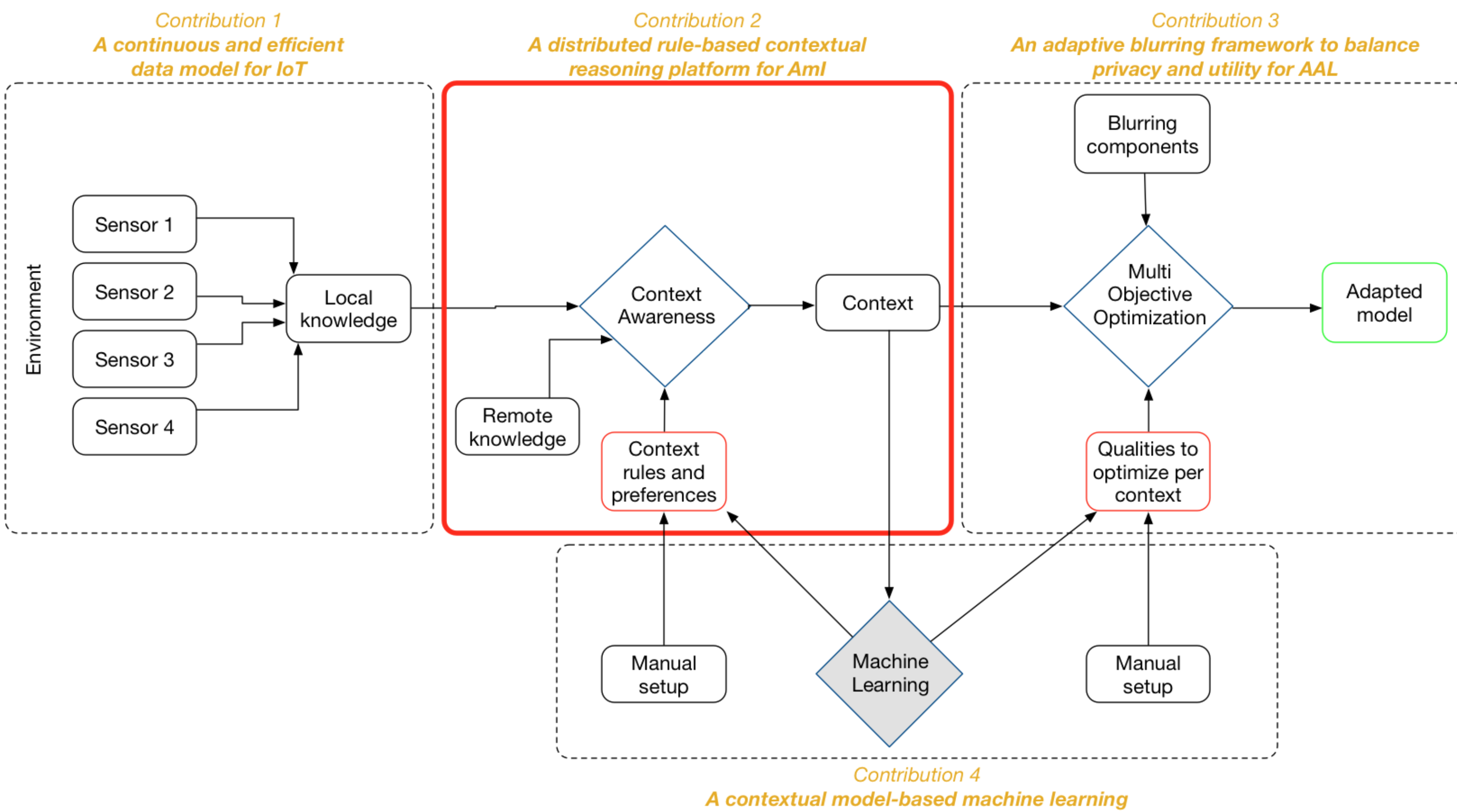


Publication:

- Assaad Moawad, Thomas Hartmann, François Fouquet, Grégory Nain, Jacques Klein, and Yves Le Traon. **Beyond discrete modeling: A continuous and efficient model for IoT**. In 2015 ACM/IEEE, 18th International Conference on Model Driven Engineering Languages and Systems (***MODELS***), Ottawa Canada, pages 90–99.

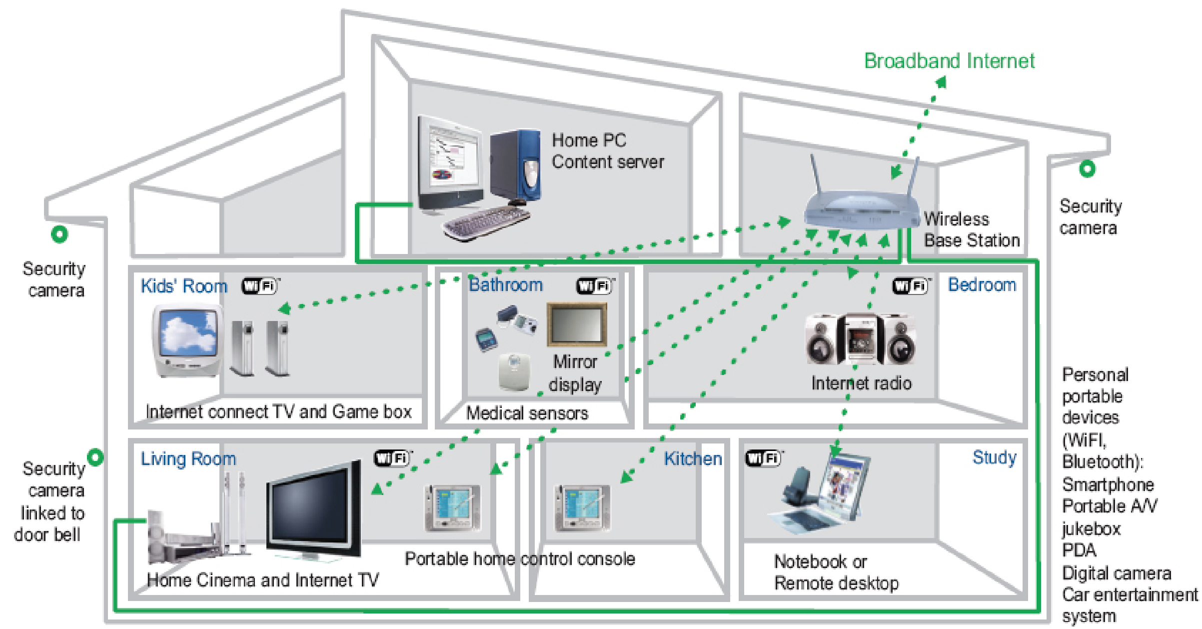
Contribution 2

A distributed context awareness for AmI



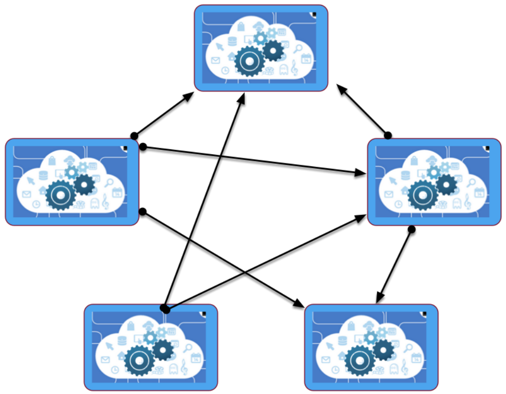
Challenges

- **Imperfect / unreliable** information
- Highly **dynamic** and open environments
- **Distributed** processing
- *How to derive* ***context?***



Multi-agent systems

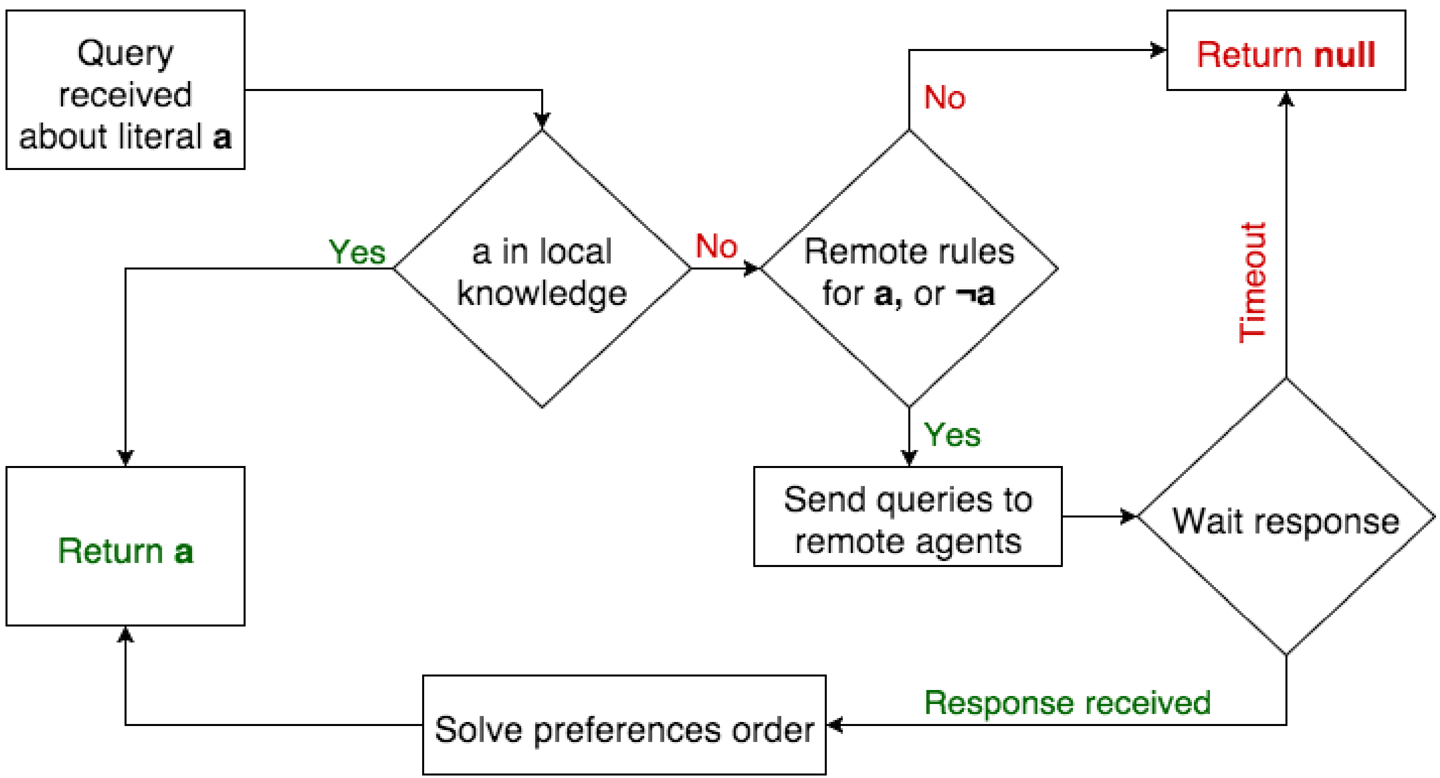
- Composed of multiple interacting **agents**
- We consider each agent to have a minimal:
  - **Computation** capabilities
  - **Communication** capabilities to other agents
  - **Local** knowledge base
  - Knowledge base about **remote** agents



Contextual defeasible logic (CDL)

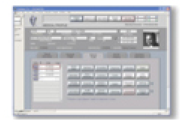
- A defeasible Multi Context system ***C***, is a collection of contexts ***Ci***
- Each Context ***Ci*** is a 3-tuple **(Vi,Ri,Ti)**:
  - **Vi**: Vocabulary used by ***Ci***. Set of logic literals (Ex: ***a***, **¬*a***)
  - **Ri**: Set of rules how to derive the literals
  - **Ti**: Preference ordering

Distributed Context awareness



Example scenario

Online medical profile



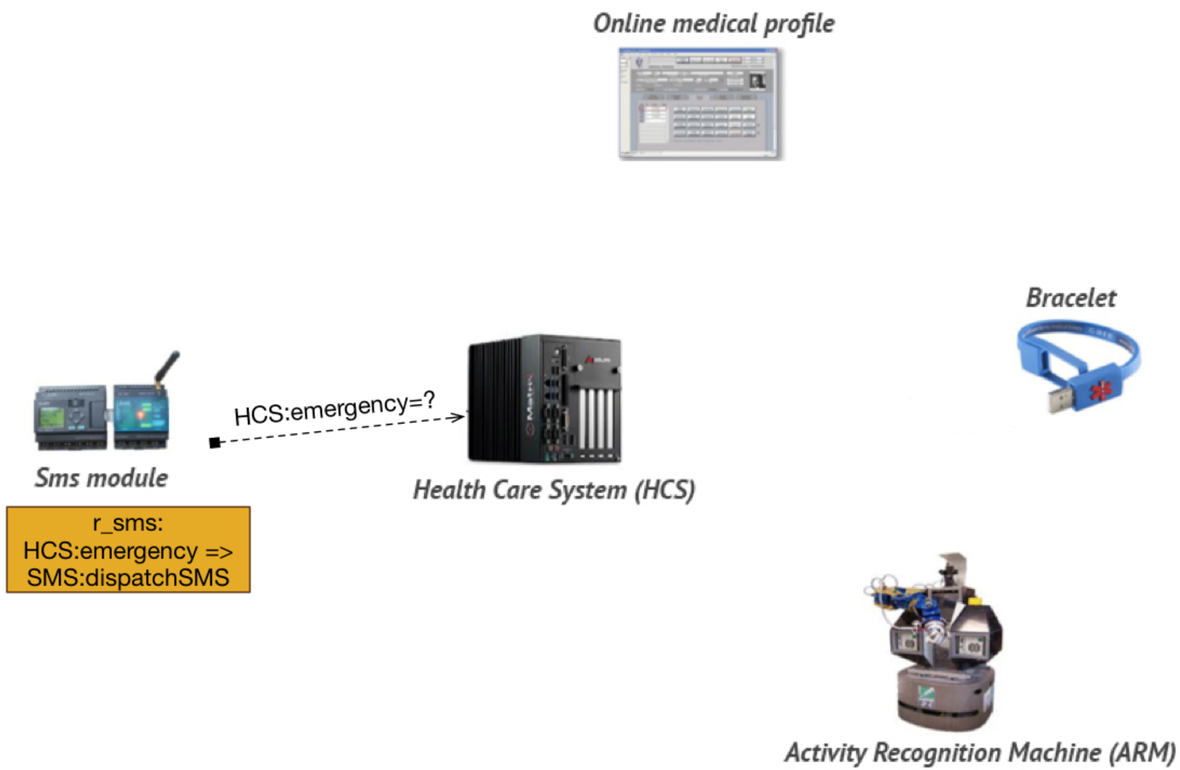
Sms module

Health Care System (HCS)

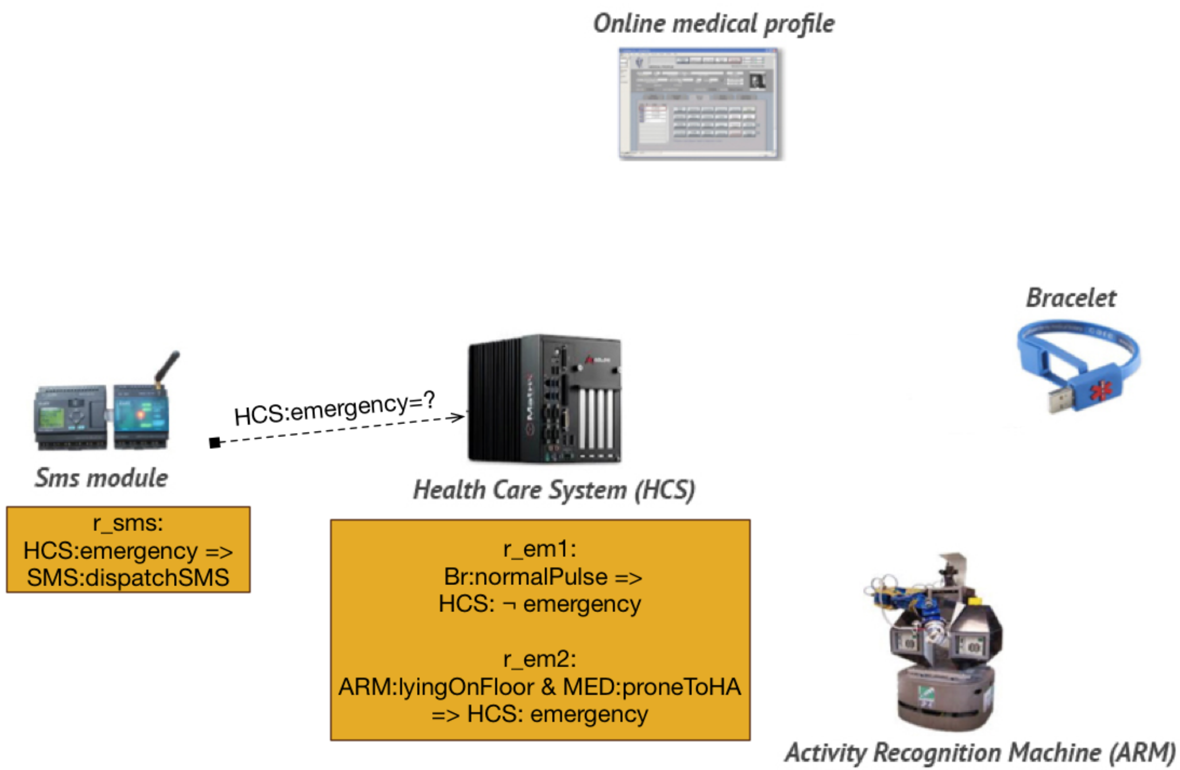
Bracelet

Activity Recognition Machine (ARM)

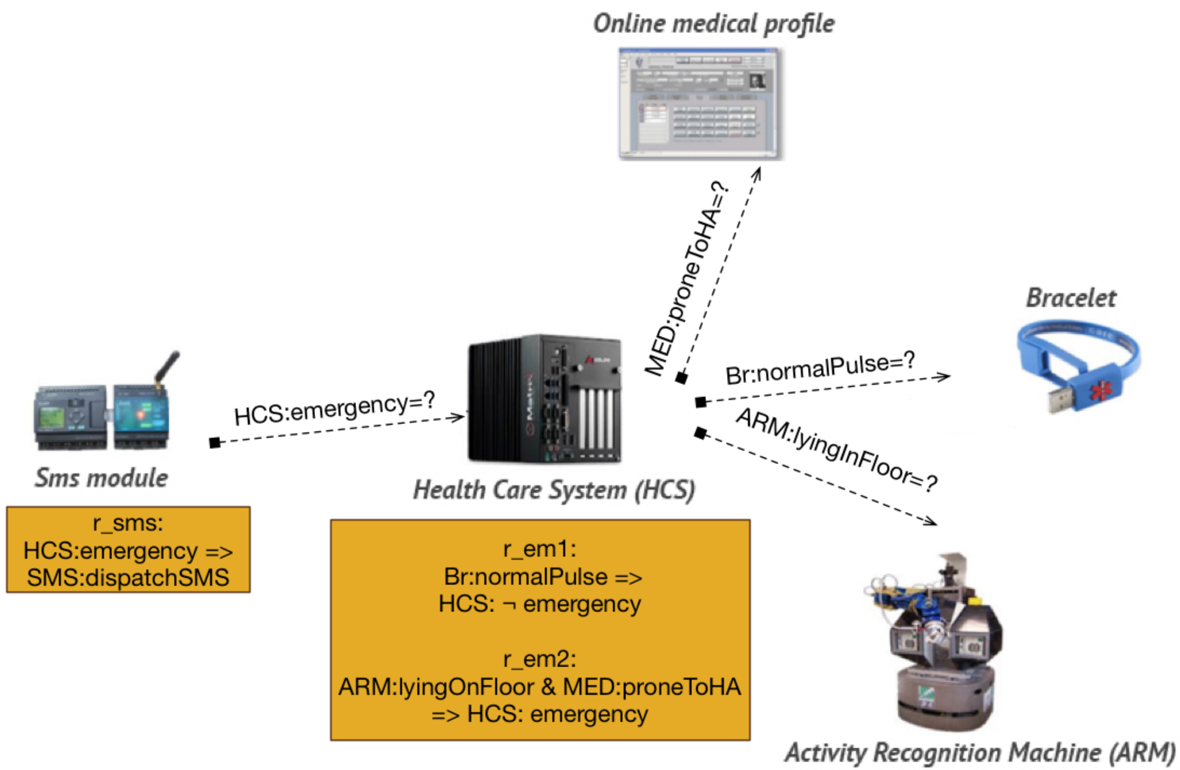
Example scenario



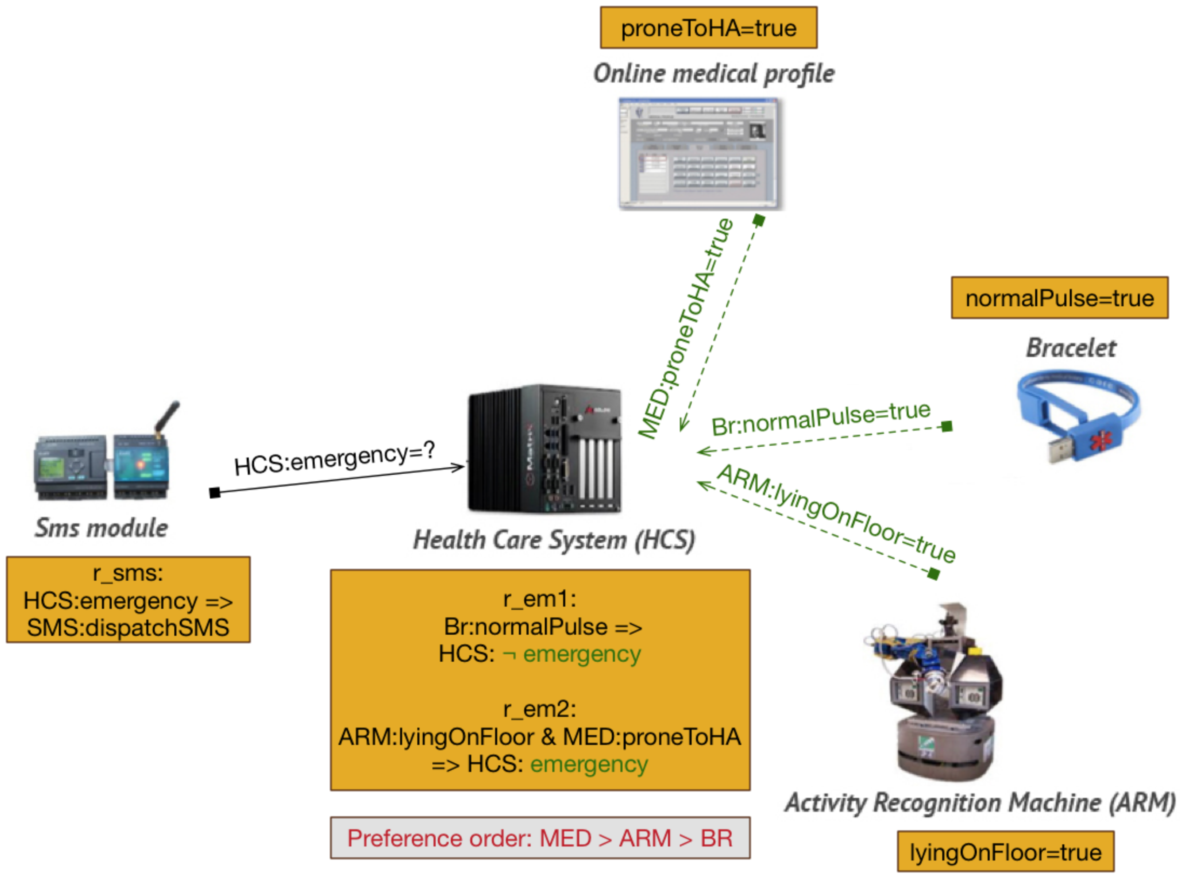
Example scenario



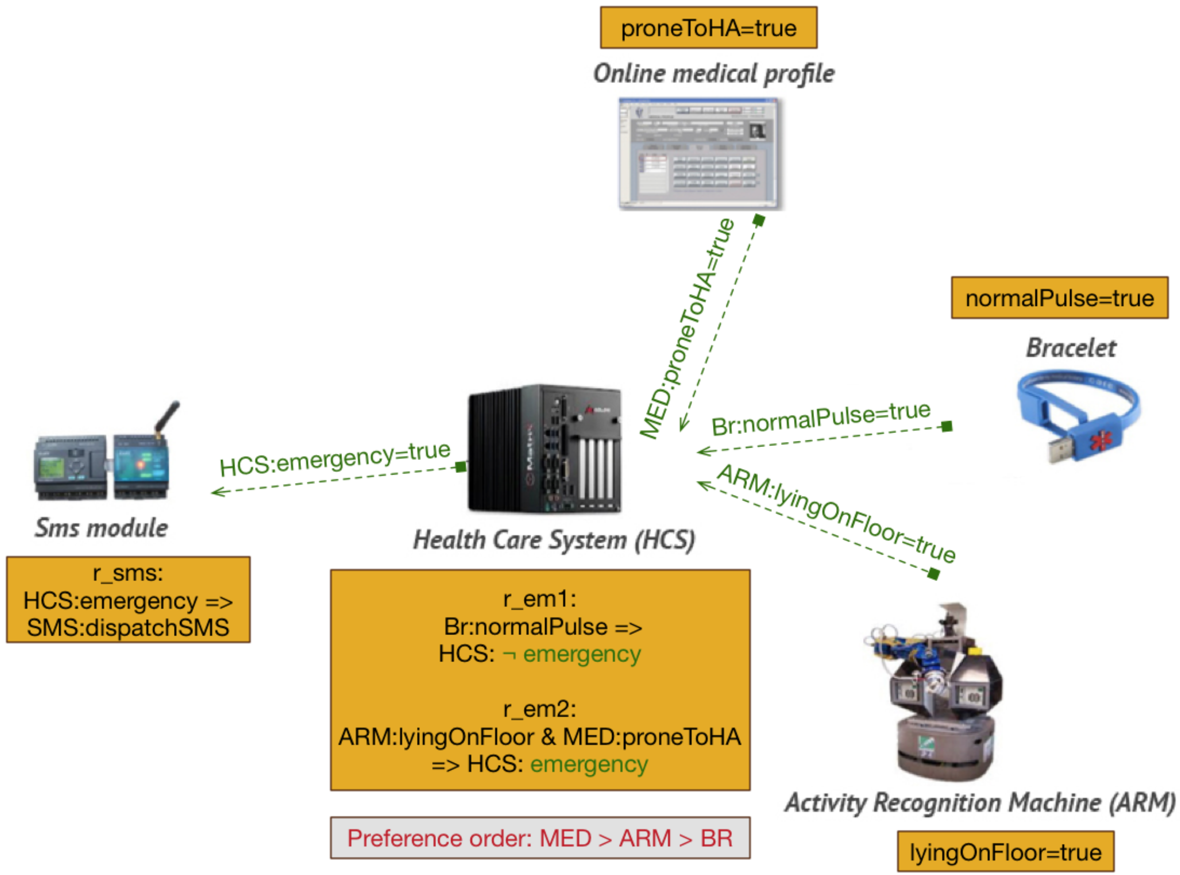
Example scenario



Example scenario



Example scenario



Deadlock problem

- By nature: ***distributed, dynamic and recursive*** processes

  -> Might cause **logic deadlocks**

- ***x_A1***->***y_A2*** and ***y_A2***->***x_A1***

  Or: ***x_A1*** -> ***y_A2*** -> ***z_A3*** -> ***t_A4*** ->… ->***x_A1***

- Cannot be detected ***a-priori***
- Loop detection at **runtime**

Deadlock problem

- By nature: ***distributed, dynamic and recursive*** processes

  -> Might cause **logic deadlocks**

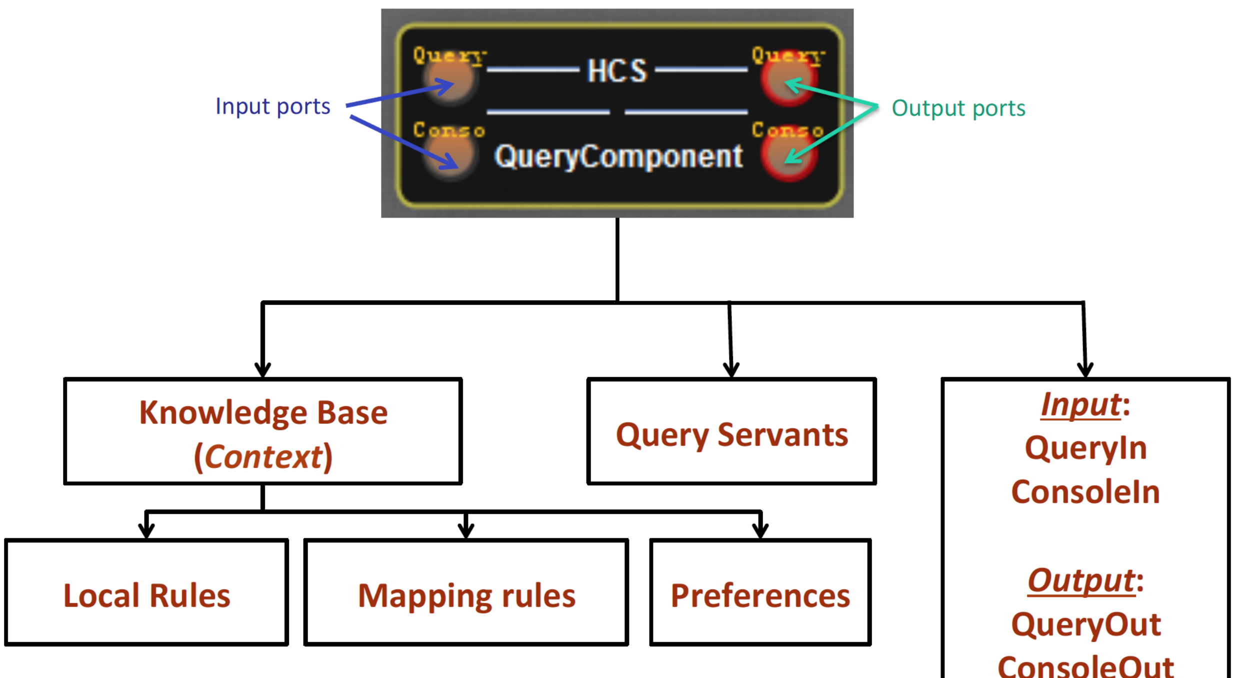
- ***x_A1***->***y_A2*** and ***y_A2***->***x_A1***

  Or: ***x_A1*** -> ***y_A2*** -> ***z_A3*** -> ***t_A4*** ->... ->***x_A1***

- **Solution:** Add history to queries to trace back the calls
- **Drawbacks:** query size & processing time increase each step

Implementation

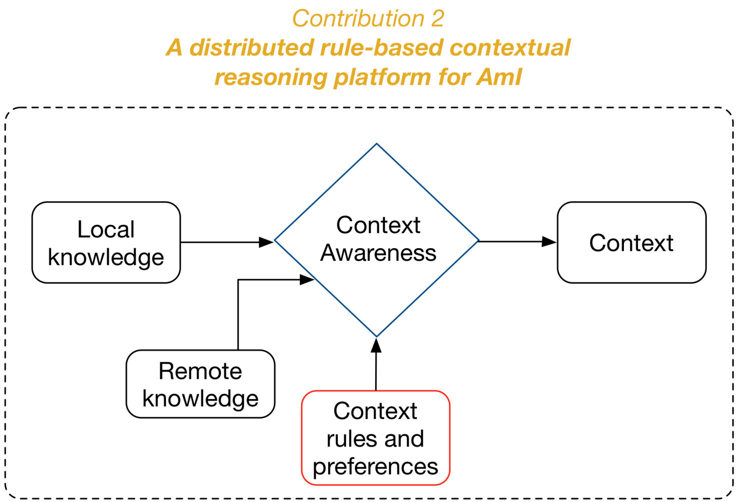
- Using **Kevoree**, *distributed component based* models:



Validation

- Tested with **500 components** with different specs/platforms
  - All queries solved correctly / loops avoided
  - Average time: ***150 ms***, interval ***[20,250] ms*** -> **Reactive**
  - Linear complexity with number of components & rules
- **Conclusion**: Fits the need of AmI & AAL

Summary

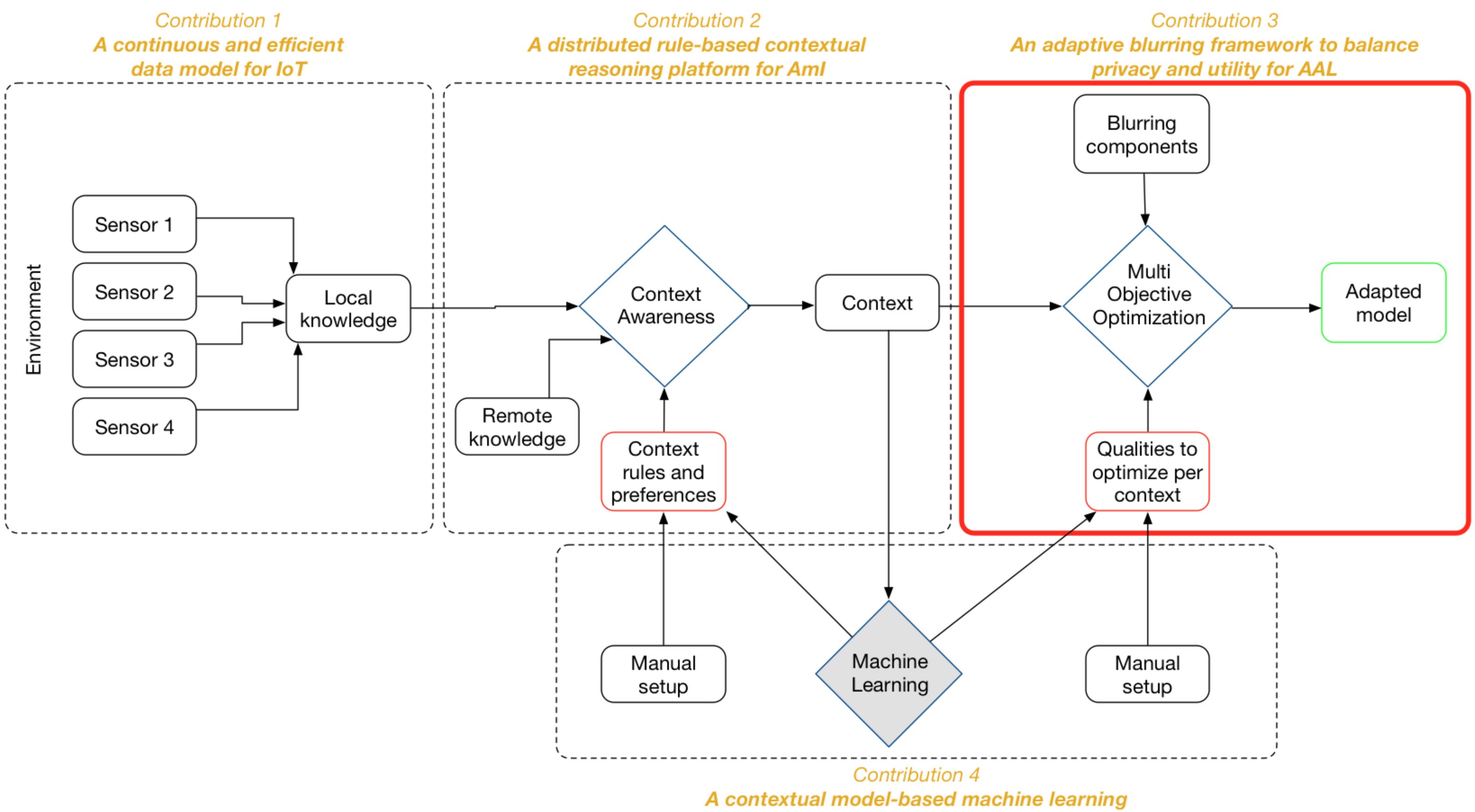


Publications:

- Assaad Moawad, Antonis Bikakis, Patrice Caire, Grégory Nain, and Yves Le Traon. **A rule-based contextual reasoning platform for ambient intelligent environments**. 7th International Symposium, ***RuleML***, seattle USA, Springer 2013, pages 158–172.
- Assaad Moawad, Antonis Bikakis, Patrice Caire, Grégory Nain, and Yves Le Traon. **Rcore: A rule-based contextual reasoning platform for AmI**. 7th International Symposium, ***RuleML@ChallengeEnriched*** demo, 2013.

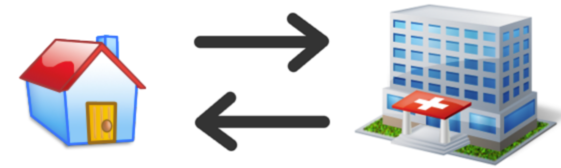
Contribution 3

An adaptive platform for AAL



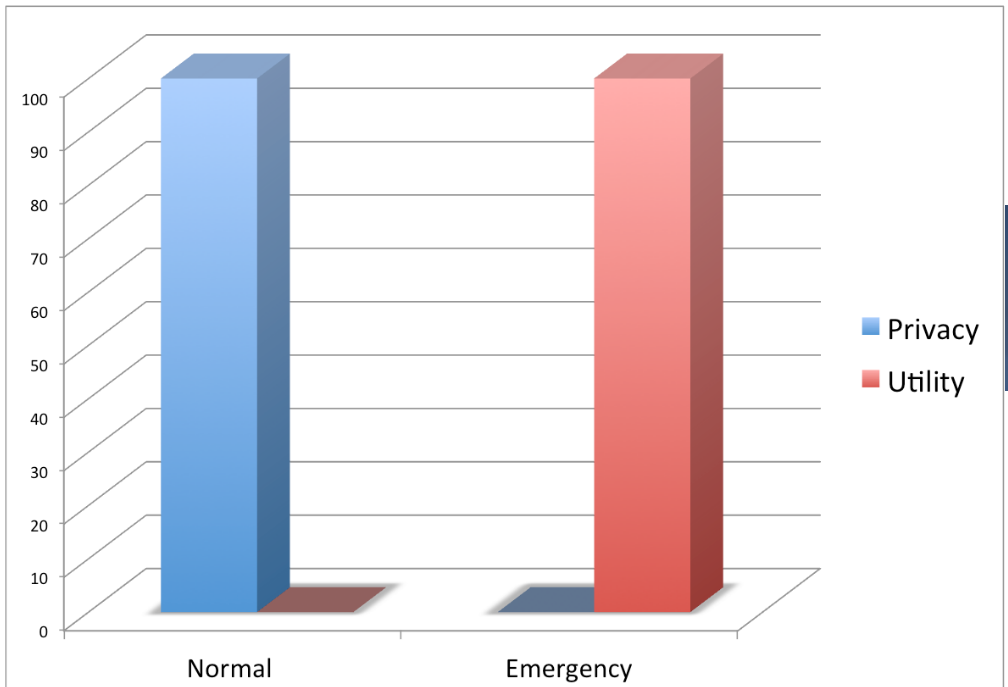
Problem statement

How to adapt the system when the ***context*** *changes?*

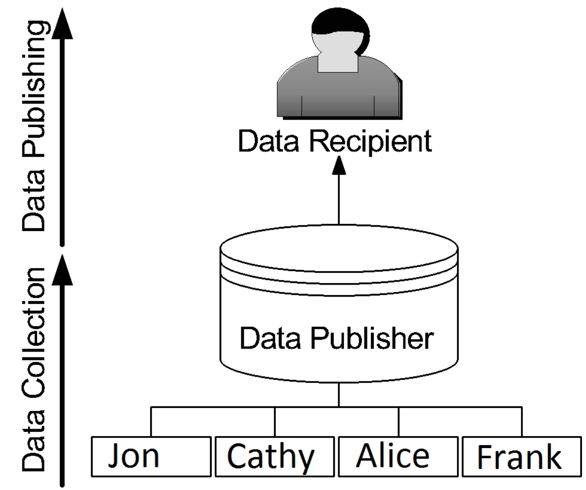


KEEP
CALM

EVERYTHING
IS UNDER CONTROL



Approaches for privacy



- Several definitions of privacy
- Several privacy risks
- Most known techniques/metrics:
    - K-anonymity
    - l-diversity
    - t-closeness
- **Aim**: A user is indistinguishable among ***k/l/t*** users
- Executed by the data publisher -> Not suitable for a **distributed** system

Utility

> *Quantity of information or quality of services* ***received*** *after an exchange of information*

- **Metrics**:
  - Information theory
  - Monetary value
  - User satisfaction/evaluation of a service

Binary data access

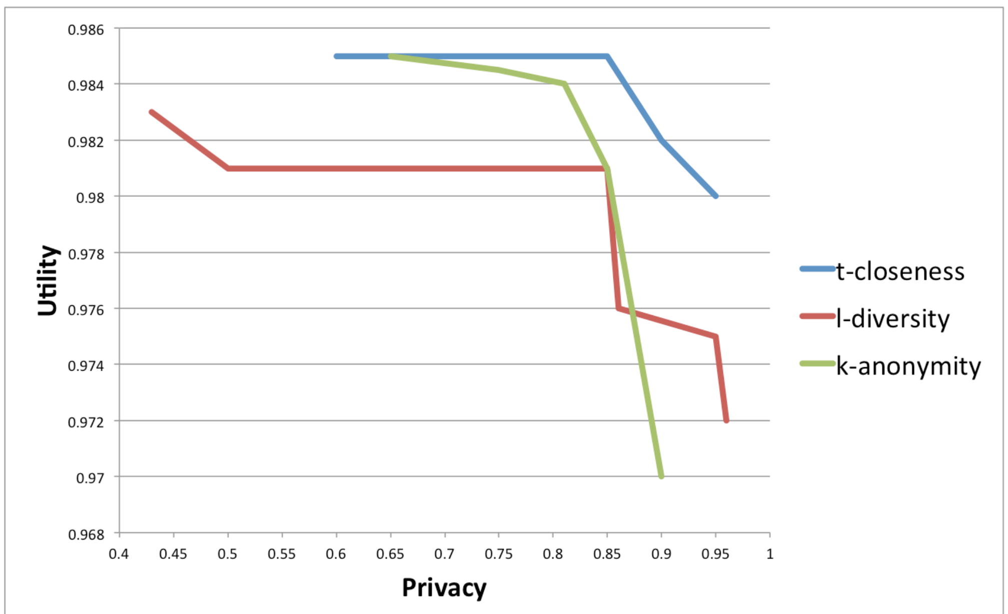
> *Binary access control (all or nothing) is not suitable for everything.*

- Sharing a precise GPS location with a weather app -> **privacy breaches**
- At the same time, sharing nothing -> **no utility**
- A region or city precision level can be a good **trade-off**

Privacy vs utility

Sending more information does not necessary increase the utility received.

The trade-off between privacy-utility is ***not linear***

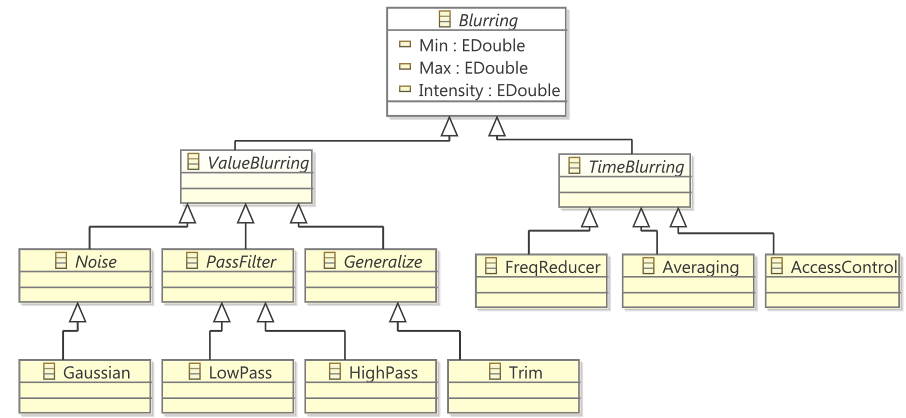


- Ex. Electric consumption:
  104.56766 W/h
- Electric consumption:
  100-200 W/h

How to share only necessary information?

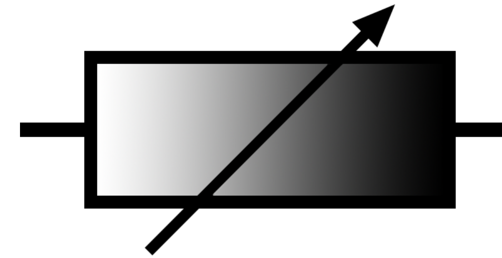
Blurring components

- **Value blurring:**
  - Noise: 2.345 -> 2.5247
  - Generalizing: 2.365 -> 2
- **Time blurring:**
  - Frequency reducing (1/sec -> 1/min)
  - Averaging over a period of time
  - Forbid access in certain periods

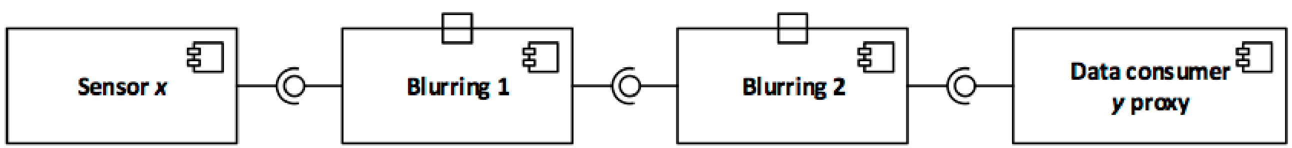


Proportional data access

> *Blurring components offer a proportional data access*

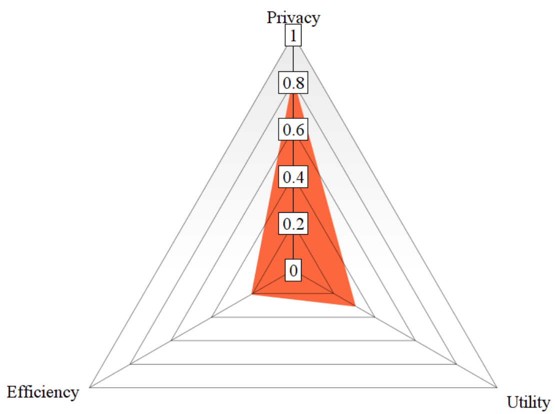


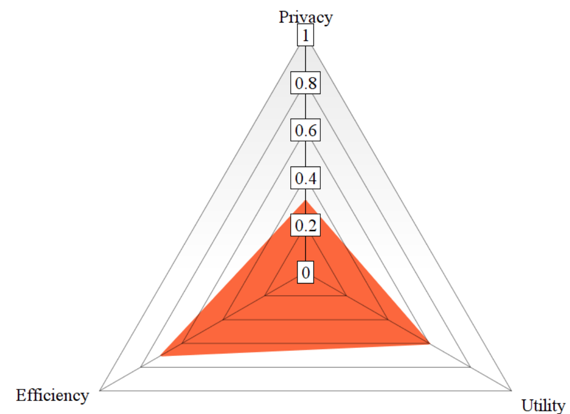
- Can have a **variable** intensity
- Can be **cascaded** to form a chain
- **Efficiency drawback**, for ex: blurring a video stream



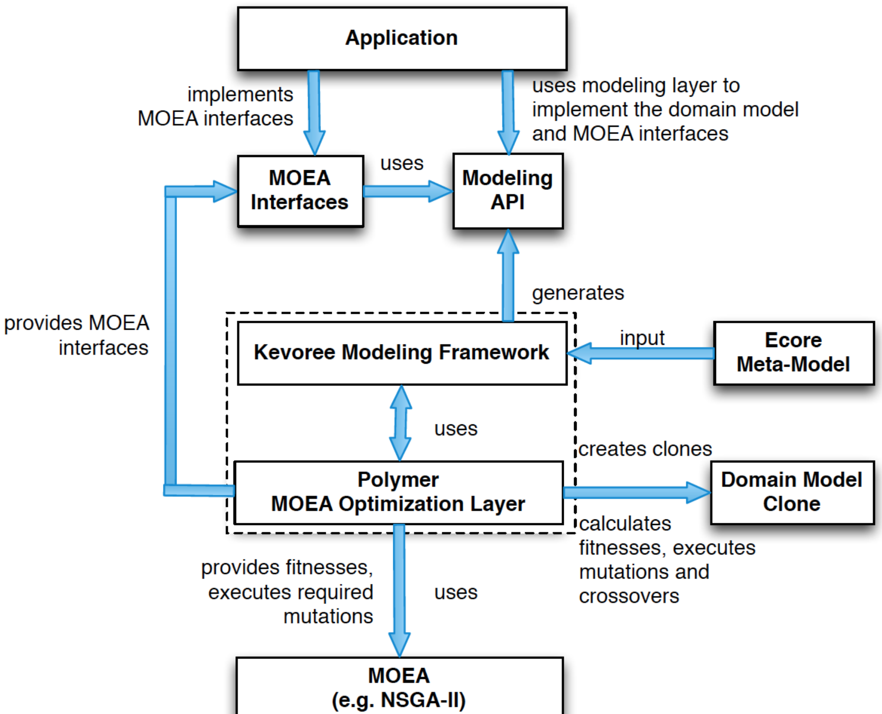
Finding a trade-off

- Several **conflicting objectives** to optimize
- How to find the good blurring chain and its parameters ?
- ***Solution:*** Multi Objective Evolutionary Algorithms (***MOEA***)
- But first, how to run MOEAs on top of **component models** ?





Polymer framework

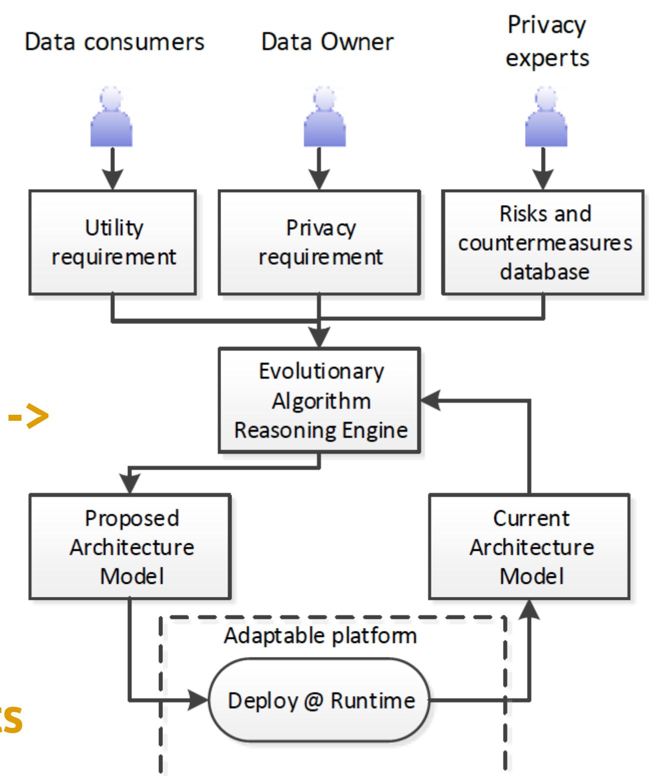


Generic contribution: allows to execute MOEA on top of **models** generated by ***KMF/Kevoree***

Publication:

- Assaad Moawad, Thomas Hartmann, François Fouquet, Grégory Nain, Jacques Klein, and Johann Bourcier. **Polymer: A model-driven approach for simpler, safer, and evolutive multi-objective optimization development**. In ***MODELSWARD 2015***, pages 286–293.

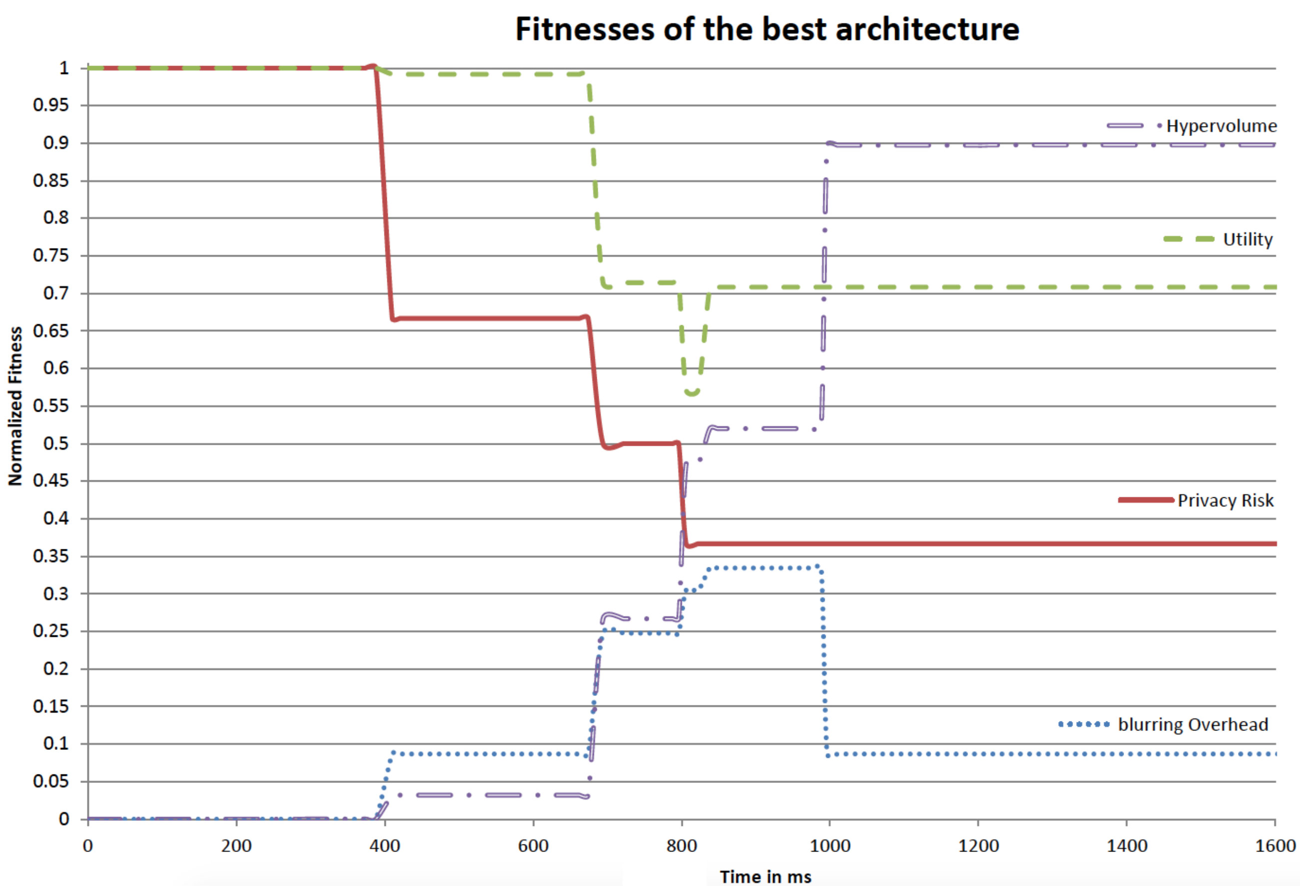
Adaptive blurring framework

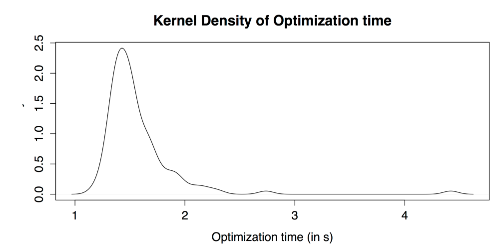


Polymer framework ->

Kevoree components

Execution

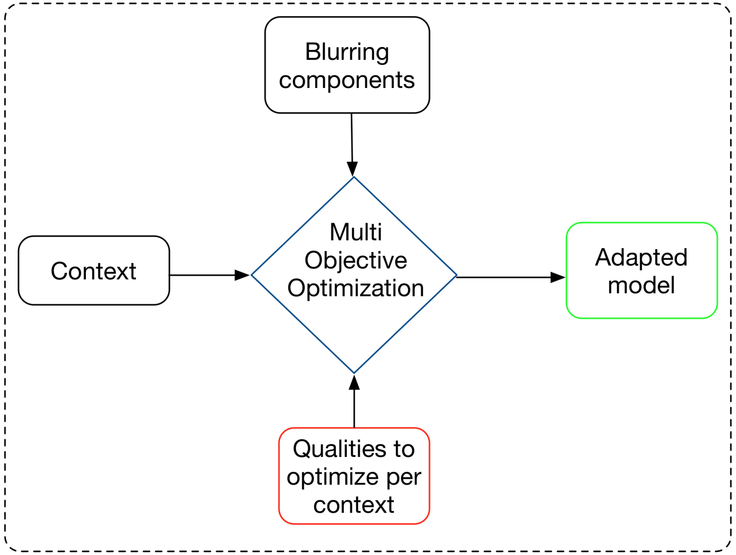

Fitnesses of the best architecture
Normalized Fitness
Time in ms
Hypervolume
Utility
Privacy Risk
blurring Overhead
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
200
400
600
800
1000
1200
1400
1600


Kernel Density of Optimization time
Optimization time (in s)
0.0
0.5
1.0
1.5
2.0
2.5
1
2
3
4

Summary

Contribution 3

An adaptive blurring framework to balance privacy and utility for AAL

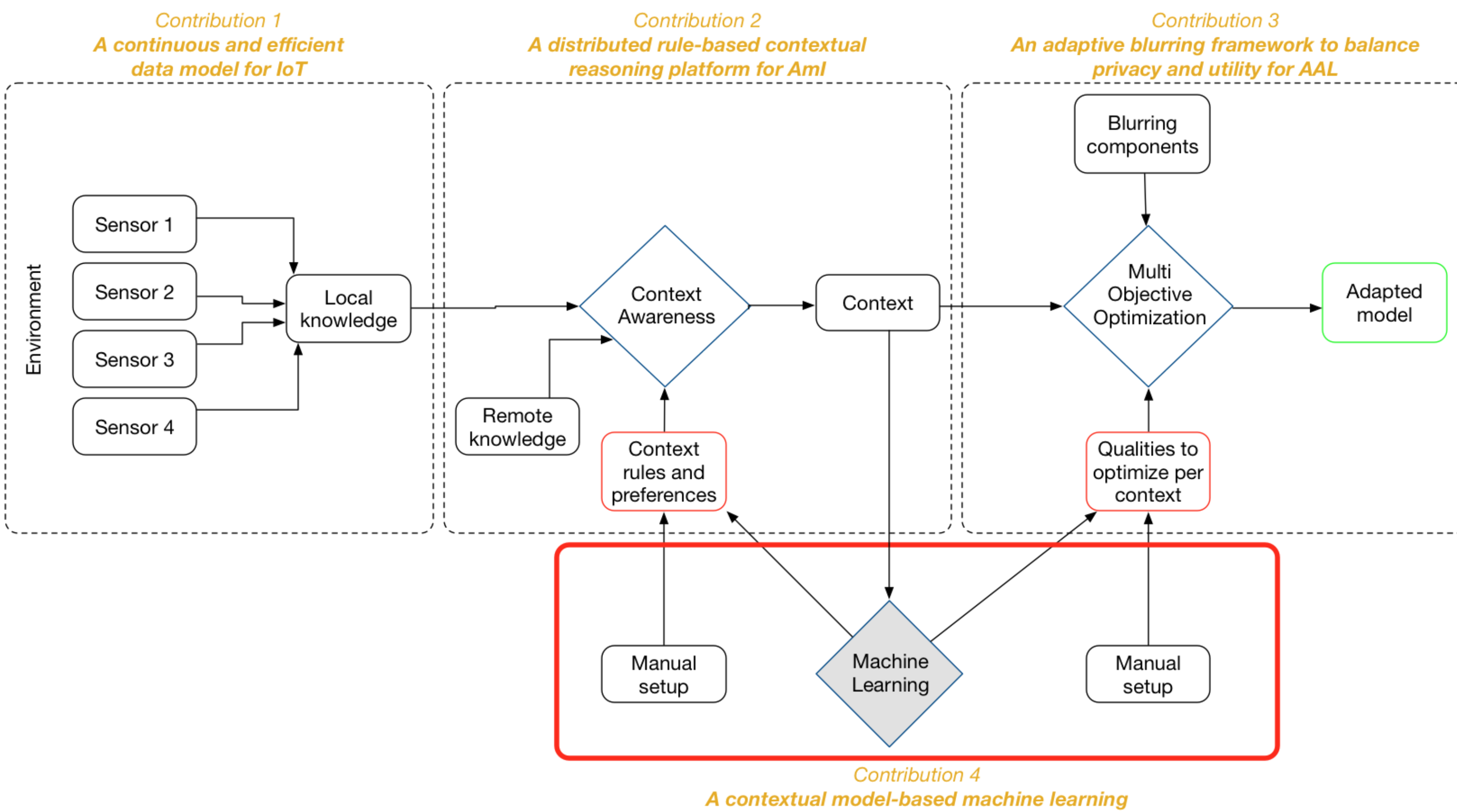


Publications:

- Patrice Caire, Assaad Moawad, Vasilis Efthymiou, Antonis Bikakis, and Yves Le Traon. **Privacy challenges in Ambient intelligent systems**. Journal of Ambient Intelligence and Smart Environments (*JAISE*). Accepted.
- Assaad Moawad, Thomas Hartmann, François Fouquet, Grégory Nain, Jacques Klein, and Johann Bourcier. **Polymer: A model-driven approach for simpler, safer, and evolutive multi-objective optimization development**. In ***MODELSWARD 2015***, pages 286−293.
- Assaad Moawad, Thomas Hartmann, François Fouquet, Jacques Klein, and Yves Le Traon. **Adaptive blurring of sensor data to balance privacy and utility for ubiquitous services.** In SAC 2015, the 30th ACM/SIGAPP Symposium On Applied Computing, pages 2271−2278. ACM, 2015

Contribution 4

Contextual model-based machine learning



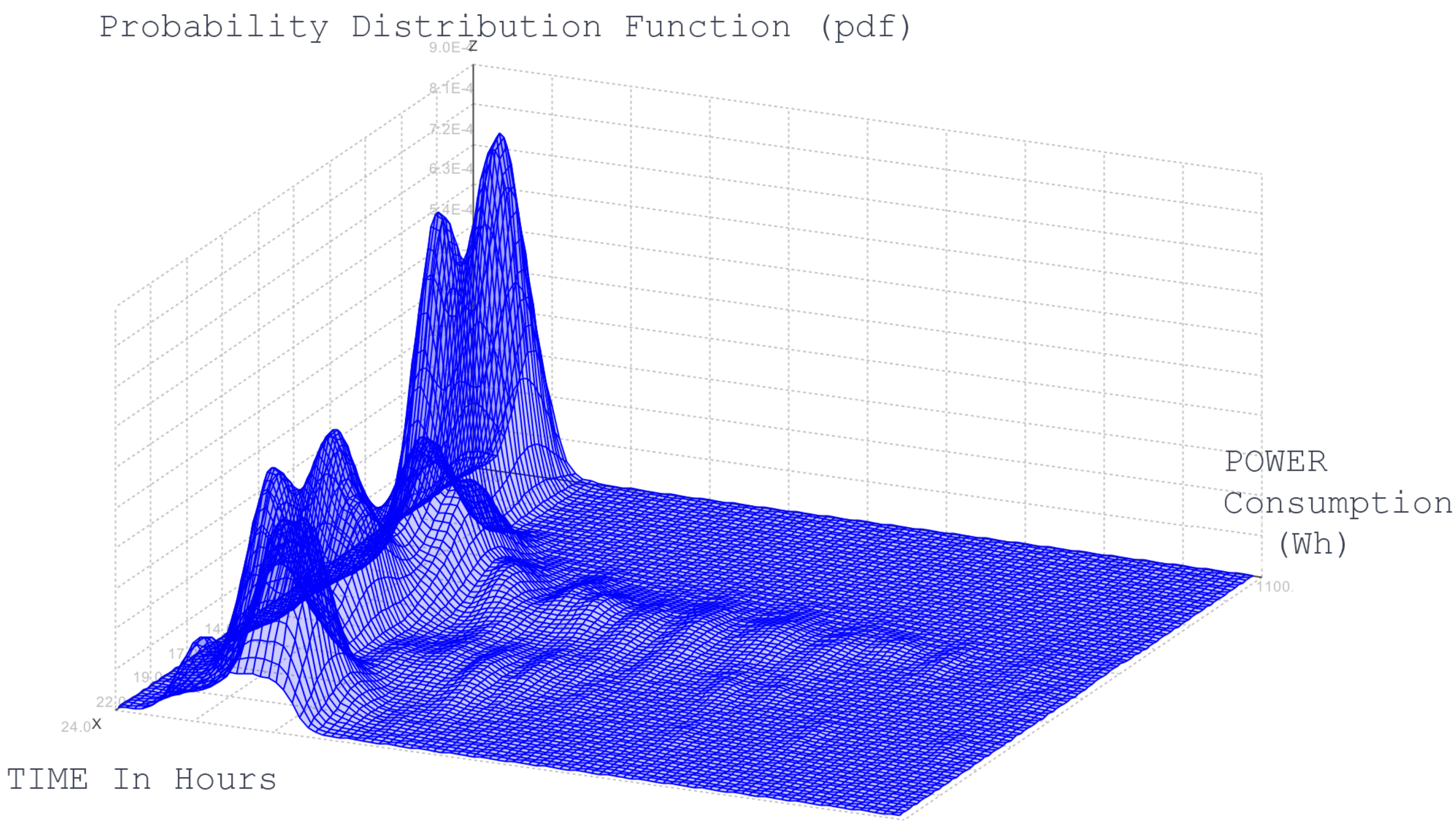
Research questions

- How to get benefits from contextual information?
- How to **detect contexts** automatically? (***Ongoing work***)

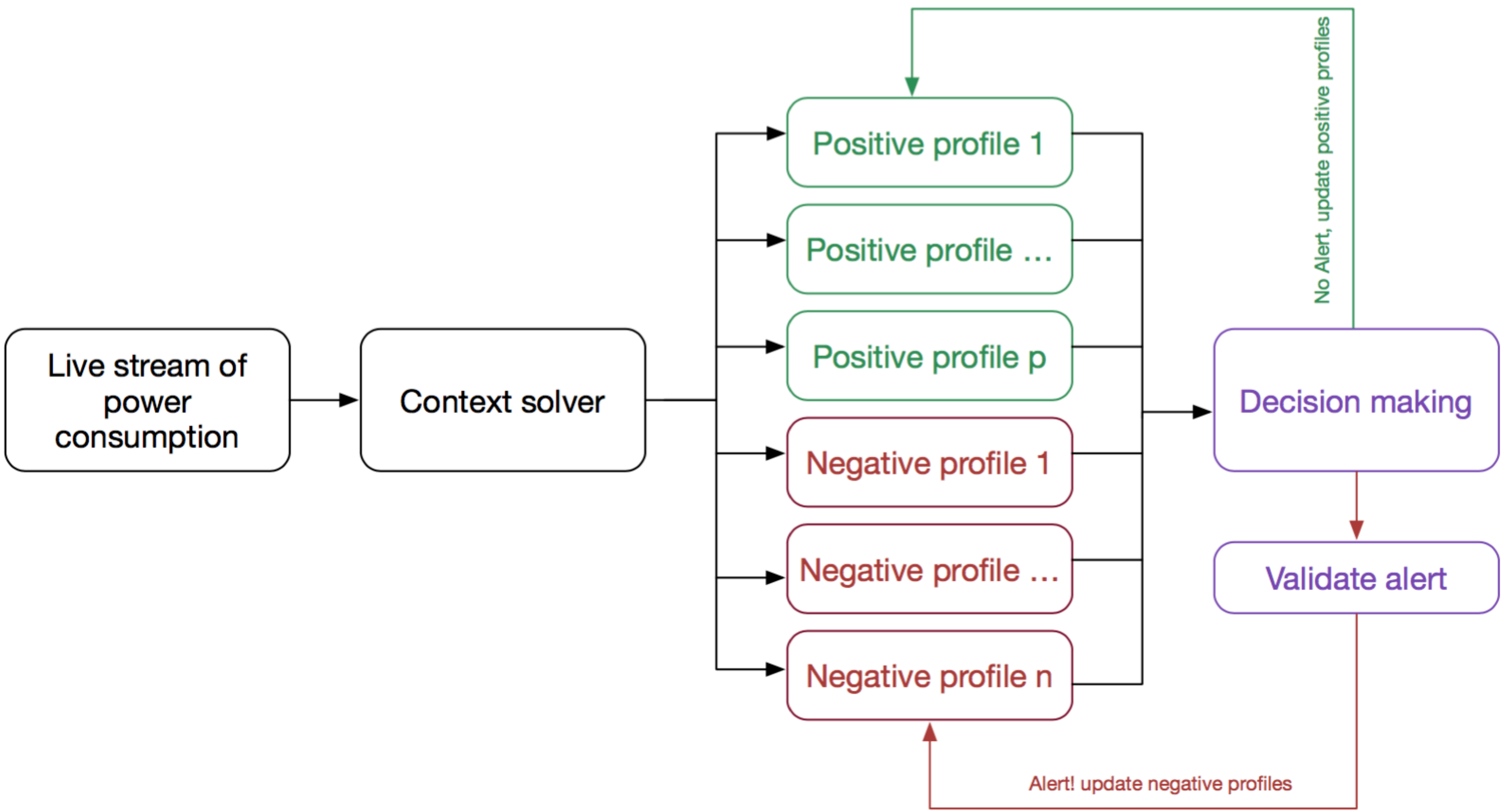
First application domain

- Anomaly detection in electric consumption

Profiling normal behavior



Multi-Context profiling



Context-aware machine learning

- Context information -> can **improve** machine learning techniques
- Fast training: 1.37 ms/value in average
- Better results than a single profile

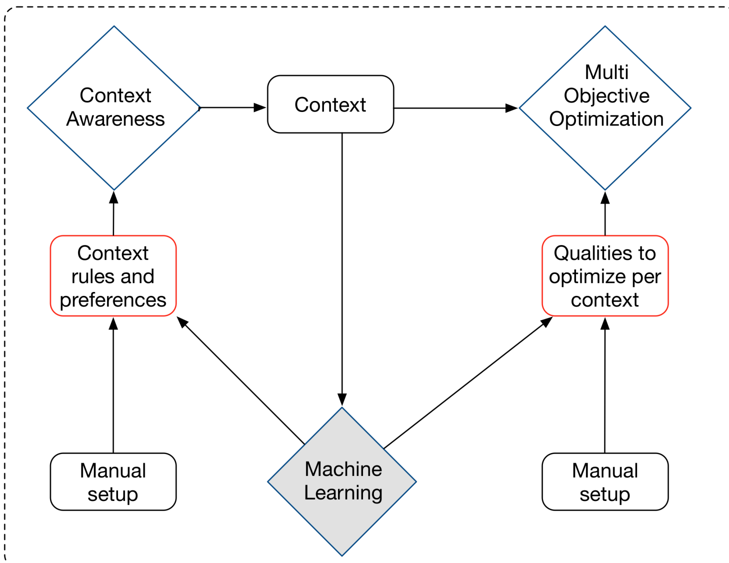
Attribute	**Single Profiler**	**Multi-context profiler**
Precision	0.602	0.808
Recall	0.99	0.99
Accuracy	0.779	0.918
F1 score	0.749	0.890

A GLOBAL OVERVEW OF RESULTS

Summary

Contribution 4

A contextual model-based machine learning

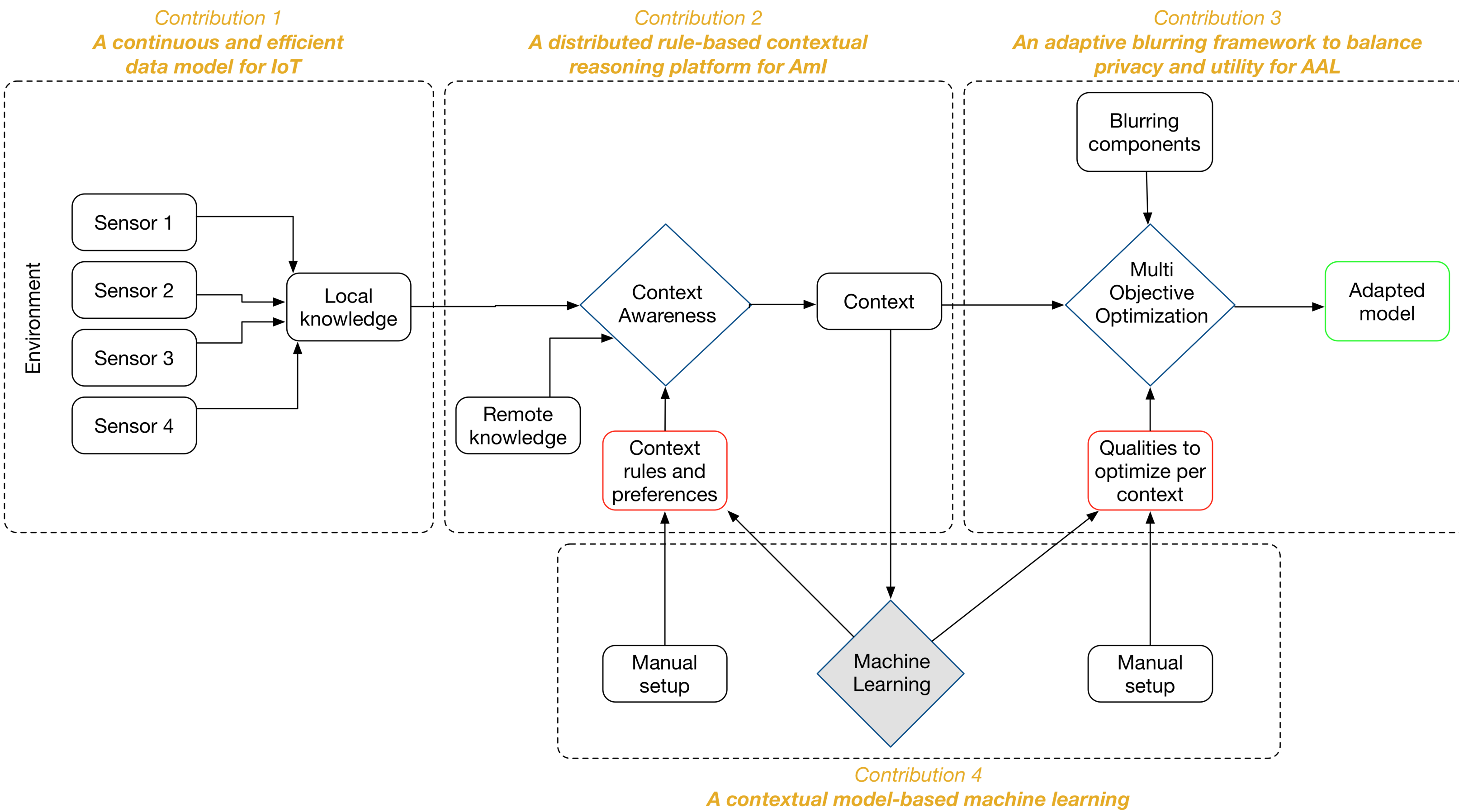


Publications:

- Thomas Hartmann, Assaad Moawad, Francois Fouquet, Yves Reckinger, Tejeddine Mouelhi, Jacques Klein, and Yves Le Traon. **Suspicious electric consumption detection based on multi-profiling using live machine learning.** In Smart Grid Communications (***SmartGridComm***), 2015 IEEE International Conference on. IEEE, 2015

Conclusion

Conclusion


Contribution 1
A continuous and efficient data model for IoT
Contribution 2
A distributed rule-based contextual reasoning platform for AmI
Contribution 3
An adaptive blurring framework to balance privacy and utility for AAL
Environment
Sensor 1
Sensor 2
Sensor 3
Sensor 4
Local knowledge
Context Awareness
Remote knowledge
Context rules and preferences
Context
Blurring components
Multi Objective Optimization
Adapted model
Qualities to optimize per context
Manual setup
Machine Learning
Manual setup
Contribution 4
A contextual model-based machine learning

Future work

- Integrating **live machine learning** techniques into modeling tools:
  - Extend modeling DSL to express learning behavior
  - Seamless integration of RAW and learned data into the same model
- Meta-learning using **MOEA** in live
  - Optimize the learning parameters
- Adapt MOEA to work on top of data stream

THANKS YOU !

QUESTIONS ?

C1- Defining continuous meta attribute

- A continuous attribute value is define as a **sequence of weights**

$$c_{ij} = \{\ldots, w_{ijk}, \ldots\}$$

- Following the following ***formula***, these weights descripe a **polynom**:

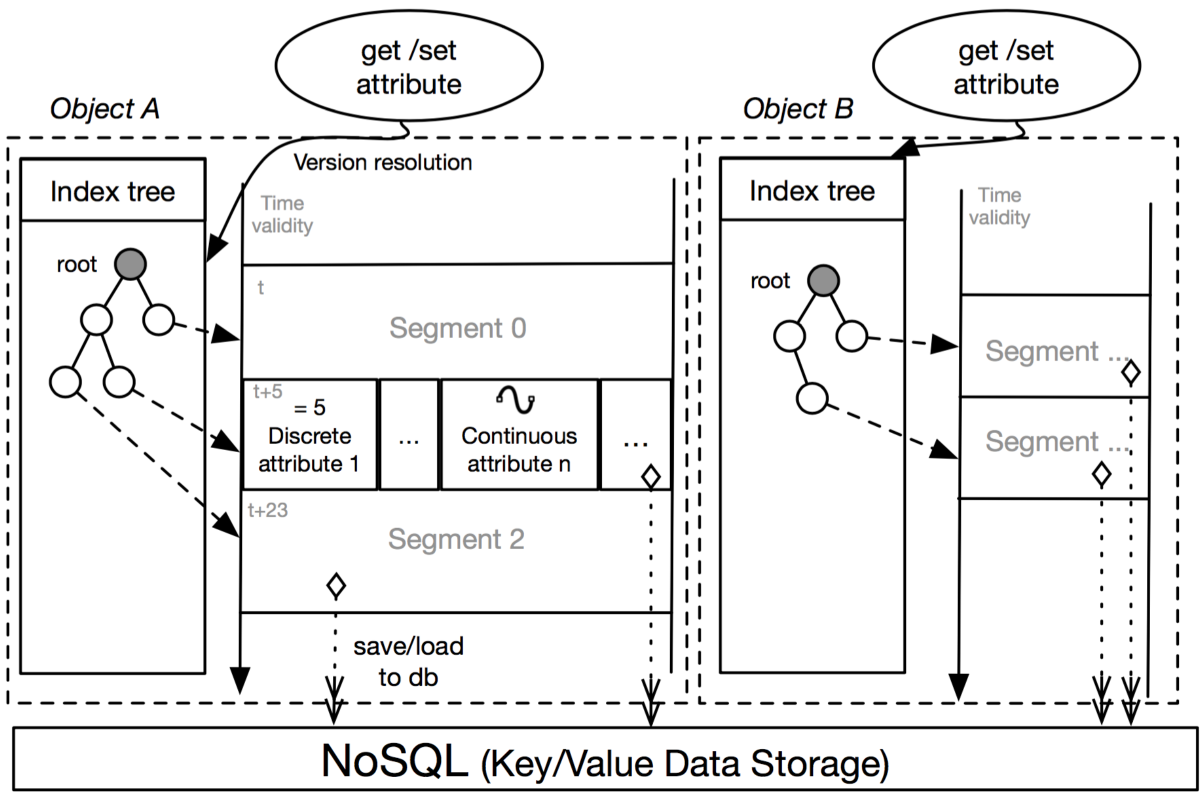
$$f_{ij}(t) = w_{ij0} + w_{ij1}(t - t_{oi}) + \ldots + w_{ijn}(t - t_{oi})^n$$

- Basic constrain:

$$\forall j, |f_{c_{ij}}(t) - y_{c_{ij}}(t)| < \epsilon_{c_{ij}}.$$

where $y_{c_{ij}}(t)$ is the physical measured value of the attribute c_{ij} at time t, and $\epsilon_{c_{ij}}$ the maximum tolerated error of this attribute as defined in the meta model.

C1- Data model structure - KMF

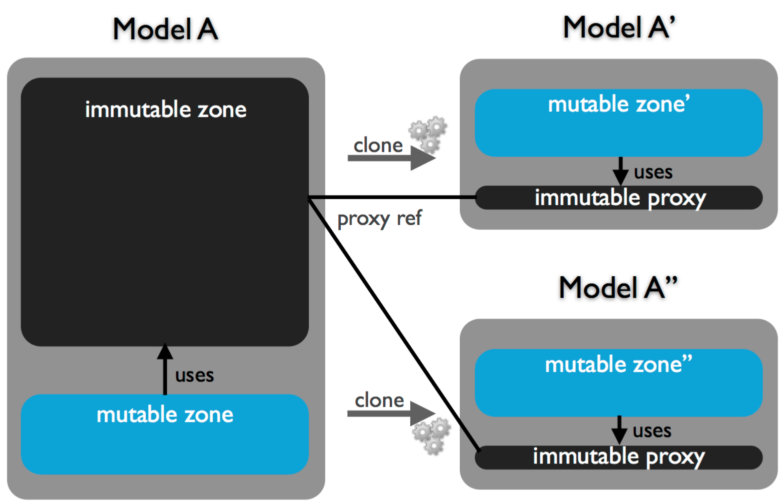


C3- The problem of encoding

- **Classical** MOEA encoding: arrays, matrices, graphs, permutations
- Encoding doesn't reflect any **semantic** or any type
- All operators need to be **manually** adapted when the encoding changes
- Skip genetic encoding -> Use **model** encoding

C3- Model-encoding problem

- A full *array copy* of genetic encoding is *cheap* for classical approach
- *Problem 2*: A full domain *model clone* can be **very expensive**
- *Solution* : **partial** clone (mutable and non mutable fields)



Integrating ML in KMF

```
class smartgrid.SmartMeter{
att activeEnergyConsumed: Double
rel profiler: smartgrid.ConsumptionProfiler
rel classifier: smartgrid.ConsumptionClassification
}
```

```
class smartgrid.ConsumptionProfiler {
with inference "GaussianProfiler" with temporalResolution 2592000000
dependency smartmeter: smartgrid.SmartMeter
input timeValue "@smartmeter | =HOURS(TIME)"
input activeEnergyConsumedValue "@smartmeter | =activeEnergyConsumed"
output probability: Double
}
```