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 Aml
Application domains

- **IoT**: Distributed, dynamic, heterogeneous
- **AMI**: User-centric, adaptive, unobtrusive
- **AAL**: Private, reactive, cost-effective
Background - frameworks

- **IoT**: Distributed, dynamic and heterogeneous

Models@run.time  
Component-based middleware

**Kevoree Modeling Framework**

*Free the code from models!*
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
A continuous and efficient data model for IoT

Contribution 2
A distributed rule-based contextual reasoning platform for Aml

Contribution 3
An adaptive blurring framework to balance privacy and utility for AAL

Contribution 4
A contextual model-based machine learning
Reasoning on physical measurements

- 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)

"**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
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.

<table>
<thead>
<tr>
<th>Database</th>
<th>Sensor</th>
</tr>
</thead>
<tbody>
<tr>
<td>DS1: Constant</td>
<td>c=42</td>
</tr>
<tr>
<td>DS2: Linear function</td>
<td>y=5x</td>
</tr>
<tr>
<td>DS3: Temperature</td>
<td>DHT11 (0 50°C +/- 2°C)</td>
</tr>
<tr>
<td>DS4: Luminosity</td>
<td>SEN-09088 (10 lux precision)</td>
</tr>
<tr>
<td>DS5: Electricity load</td>
<td>from Creos SmartMeters data</td>
</tr>
<tr>
<td>DS6: Music file</td>
<td>2 minutes samples from wav file</td>
</tr>
<tr>
<td>DS7: Pure random</td>
<td>from random.org</td>
</tr>
</tbody>
</table>
**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

<table>
<thead>
<tr>
<th>Database</th>
<th>Discrete</th>
<th>Polynomial</th>
</tr>
</thead>
<tbody>
<tr>
<td>DS1: Constant</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>DS2: Linear function</td>
<td>5%</td>
<td>0%</td>
</tr>
<tr>
<td>DS3: Temperature</td>
<td>8.5%</td>
<td>3%</td>
</tr>
<tr>
<td>DS4: Luminosity</td>
<td>9.9%</td>
<td>3.5%</td>
</tr>
<tr>
<td>DS5: Electricity</td>
<td>17%</td>
<td>6%</td>
</tr>
<tr>
<td>DS6: Sound sensor</td>
<td>21%</td>
<td>13%</td>
</tr>
<tr>
<td>DS7: Random</td>
<td>31.8%</td>
<td>30.8%</td>
</tr>
</tbody>
</table>

Average error when we try to approximate missing values
Summary

**Contribution 1**

*A continuous and efficient data model for IoT*

---

**Publication:**
Contribution 2
A distributed context awareness for Aml

Contribution 1
A continuous and efficient data model for IoT

Contribution 2
A distributed rule-based contextual reasoning platform for Aml

Contribution 3
An adaptive blurring framework to balance privacy and utility for AAL

Contribution 4
A contextual model-based machine learning

Environment
- Sensor 1
- Sensor 2
- Sensor 3
- Sensor 4

Local knowledge

Context Awareness

Remote knowledge

Context rules and preferences

Context

Multi-Objective Optimization

Blurring components

Adapted model

Manual setup

Machine Learning

Manual setup

Qualities to optimize per context
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 $C_i$
- Each Context $C_i$ is a 3-tuple $(V_i, R_i, T_i)$:
  - $V_i$: Vocabulary used by $C_i$. Set of logic literals (Ex: $a$, $\neg a$)
  - $R_i$: Set of rules how to derive the literals
  - $T_i$: Preference ordering
Distributed Context awareness

Query received about literal a

- Yes: a in local knowledge
  - Yes: Send queries to remote agents
    - Yes: Wait response
    - No: Remote rules for a, or ¬a
      - No: Return null
      - Yes: Return a
  - No: Solve preferences order

- No: Solve preferences order
  - Yes: Wait response
    - Yes: Response received
    - No: Timeout

Return null
Example scenario

Online medical profile

Sms module

Health Care System (HCS)

Bracelet

Activity Recognition Machine (ARM)
Example scenario
Example scenario

- **SMS module**
  - r_sms:
    - HCS: emergency => SMS: dispatchSMS

- **Health Care System (HCS)**
  - r_em1:
    - Br: normalPulse => HCS: non-emergency
  - r_em2:
    - ARM: lyingOnFloor & MED: proneToHA
      => HCS: emergency

- **Online medical profile**

- **Bracelet**

- **Activity Recognition Machine (ARM)**
Example scenario

- **Online medical profile**
- **Health Care System (HCS)**
- **Activity Recognition Machine (ARM)**

**Rules:**
- r_sms:
  - HCS:emergency => SMS:dispatchSMS
- r_em1:
  - Br:nominalPulse => HCS: ¬ emergency
- r_em2:
  - ARM:lyingOnFloor & MED:proneToHA => HCS: emergency
Example scenario

Health Care System (HCS)

r_sms:
HCS:emergency =>
SMS:dispatchSMS

r_em1:
Br:normalPulse =>
HCS: ~ emergency

r_em2:
ARM:lyingOnFloor & MED:proneToHA => HCS: emergency

Preference order: MED > ARM > BR

Online medical profile

proneToHA=true
normalPulse=true

Br: normalPulse = true
ARM: lyingOnFloor = true

Sms module

Activity Recognition Machine (ARM)

lyingOnFloor = true

Bracelet
Example scenario

- **Sms module**
  - r_sms: HCS:emergency => SMS:dispatchSMS

- **Health Care System (HCS)**
  - HCS:emergency=true
  - MED:proneToHA=true

- **Activity Recognition Machine (ARM)**
  - ARM:lyingOnFloor=true

- **Bracelet**
  - Br:normalPulse=true

- **Preference order**: MED > ARM > BR

**Online medical profile**

- proneToHA=true
- normalPulse=true
Deadlock problem

- By nature: *distributed, dynamic and recursive* processes
  - Might cause *logic deadlocks*
- $x_{A1} \rightarrow y_{A2}$ and $y_{A2} \rightarrow x_{A1}$
  - Or: $x_{A1} \rightarrow y_{A2} \rightarrow z_{A3} \rightarrow t_{A4} \rightarrow \ldots \rightarrow x_{A1}$
- Cannot be detected *a-priori*
- Loop detection at *runtime*
Deadlock problem

- By nature: *distributed, dynamic and recursive* processes
  - Might cause *logic deadlocks*
- \( x_{A1} \rightarrow y_{A2} \) and \( y_{A2} \rightarrow x_{A1} \)
  - Or: \( x_{A1} \rightarrow y_{A2} \rightarrow z_{A3} \rightarrow t_{A4} \rightarrow \ldots \rightarrow x_{A1} \)
- **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 Aml & AAL
Summary

**Publications:**

Contribution 3
An adaptive platform for AAL

Contribution 1
A continuous and efficient data model for IoT

Contribution 2
A distributed rule-based contextual reasoning platform for AAI

Contribution 3
An adaptive blurring framework to balance privacy and utility for AAL

Contribution 4
A contextual model-based machine learning
Problem statement

"How to adapt the system when the context changes?"
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 necessarily 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:
Adaptive blurring framework

Polymer framework ->

Kevoree components
Execution

Fitnesses of the best architecture

Kernel Density of Optimization time

Time in ms
Summary

Contribution 3
An adaptive blurring framework to balance privacy and utility for AAL

Blurring components

Multi Objective Optimization

Adapted model

Context

Qualities to optimize per context

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.

Contribution 4
Contextual model-based machine learning

Contribution 1
A continuous and efficient data model for IoT

Contribution 2
A distributed rule-based contextual reasoning platform for AML

Contribution 3
An adaptive blurring framework to balance privacy and utility for AAL

Contribution 4
A contextual model-based machine learning

Environment

Sensor 1
Sensor 2
Sensor 3
Sensor 4

Local knowledge

Remote knowledge

Context rules and preferences

Context Awareness

Context

Multiday Objective Optimization

Blurring components

Qualities to optimize per context

Manual setup

Machine Learning

Manual setup

Adapted model
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

Live stream of power consumption → Context solver → Positive profile 1 → Positive profile … → Positive profile p → Decision making

Negative profile 1 → Negative profile … → Negative profile n → Validate alert → Alert! update negative profiles

No Alert, update positive profiles
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

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Single Profiler</th>
<th>Multi-context profiler</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>0.602</td>
<td>0.808</td>
</tr>
<tr>
<td>Recall</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.779</td>
<td>0.918</td>
</tr>
<tr>
<td>F1 score</td>
<td>0.749</td>
<td>0.890</td>
</tr>
</tbody>
</table>

**A GLOBAL OVERVIEW OF RESULTS**
Summary

Contribution 4

A contextual model-based machine learning

- **Context Awareness**
- **Context**
- **Context rules and preferences**
- **Machine Learning**
- **Qualities to optimize per context**
- **Multi-Objective Optimization**
- **Manual setup**

**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

Contribution 4
A contextual model-based machine learning

Environment

Sensor 1
Sensor 2
Sensor 3
Sensor 4

Local knowledge

Context Awareness

Context

Remote knowledge

Context rules and preferences

Manual setup

Machine Learning

Multi Objective Optimization

Qualities to optimize per context

Blurring components

Adapted model
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
C1- Defining continuous meta attribute

- A continuous attribute value is defined as a sequence of weights
  \[ c_{ij} = \{\ldots, w_{ijk}, \ldots\} \]

- Following the following formula, these weights describe 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

NoSQL (Key/Value Data Storage)
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
}