Experiments with Self-Stabilizing Distributed Data Fusion

B. Ducourthial, V. Cherfaoui

Sorbonne universités
Université de Technologie de Compiègne
UMR CNRS UTC 7253 Heudiasyc

September 2016

Université de Technologie de Compiègne

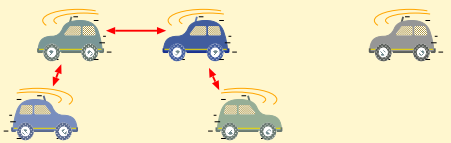
- Université de Technologie de Compiègne
  ∼4500 students, master degree (engineer diploma), PhD
  http://www.utc.fr
  - One of the first French engineering school for computer science
  - Close to Paris and Charles de Gaulle airport

- Heudiasyc lab from the UTC & CNRS
  Equipex Robotex, Labex MS2T
  https://www.hds.utc.fr

- Dynamic networks team
  https://airplug.hds.utc.fr

Agenda

1. Data fusion
2. Distributed data fusion
3. Study of the discounting **r**
4. Applications for vehicular networks
5. Conclusion



Summary

Sommaire

Data fusion

Motivation

- Several sources of information
  - How to deal with?
  - Could disagree
  - Take benefit of all of them
- Imperfect measures
  - Can we trust data?
  - Imprecision
  - Uncertainty
  - Ambiguity
- Main applications
  - Fusion of experts opinions
  - Fusion of classifiers
  - Multisensors data-fusion

Sommaire

Data fusion

Frameworks

- How to deal with imprecise and uncertain data?
  - Imprecision :
    Set Membership Approach uncertainty?
  - Aleatory uncertainty :
    Probability theory imprecision?
  - Theory of Belief Function: generalizes both
    Transferable Belief Model
    Dempster-Shafer Theory of Evidence

- Belief Function Framework
  - Information modeling
  - Combination rules

[Dempster 1968, Shafer 1976, Smets 1990s]

Sommaire

Distributed data fusion

Data representation

- Data X with value in Ω
- Representation of X
  - (value, confidence)
  - value: subset of Ω
  - confidence: indication on the reliability of the item of information
- Interest:
  - Imprecision of $X \rightsquigarrow$ value
  - Uncertainty of $X \rightsquigarrow$ confidence

[Dubois, Prade 1988]

		Confidence	
		certain	**uncertain**
Value	**precise**	20	probably 20
	imprecise	between 15 and 25	probably between 15 and 25

Sommaire

Data fusion

Fusion operators

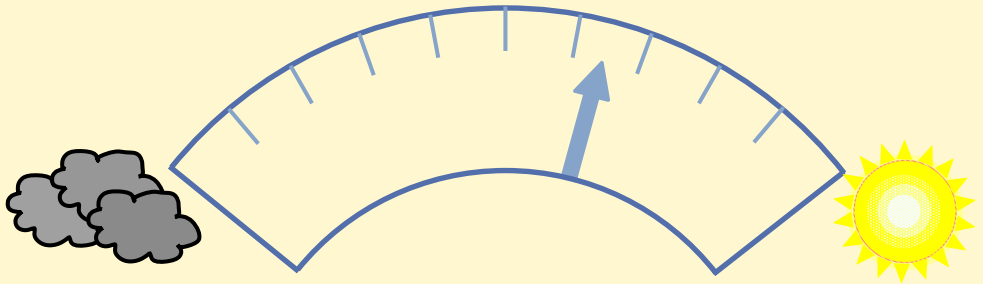
- Frame of discernment: set Ω
- Basic belief assignment
  - Mass function
    - $m^\Omega : \mathcal{P}(\Omega) \to [0,1]$
    - $\sum_{X \subset \Omega} m^\Omega(X) = 1$
  - Our algorithm: vector of weights
- Dempster operator
  - Emphases the agreement of reliable and independent sources [Smets 1990,Shafer 1976]
    $m_{1 \textcircled{\cap} 2}(A) = \sum_{B \cap C = A} m_1(B) \cdot m_2(C)$
  - Spread the conflict over other sets [Dempster]
- Cautious operator [Denoeux 2008]
  - Do not assume independent sources
  - Least commitment principle
  - Avoid the *data incest*

Sommaire

Distributed data fusion

Example of Basic Belief Assignment 1/3

- Pressure measurement

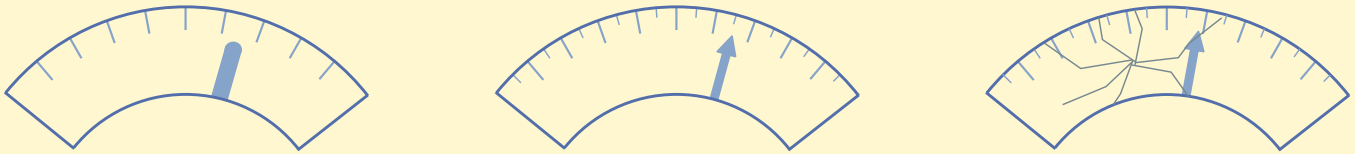


- Weather forecast
  - Compare current measure with the last one

Distributed data fusion

Example of Basic Belief Assignment 2/3

- Barometer?

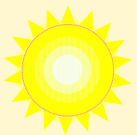


- Measure:
  - Pressure measurement: interval $I \subset \mathbb{R}^+$
  - Pressure gradient: interval $\Delta I \subset \mathbb{R}$
  - *Simple* mass function:
    - Only two subsets: ΔI and $\mathbb{R}$
    - $\mathbb{R}$: lack of knowledge
    - $m^{\mathbb{R}}(\Delta I) = 1 - \alpha$
    - $m^{\mathbb{R}}(\mathbb{R}) = \alpha$
    - α: uncertainty of the barometer

Distributed data fusion

Example of Basic Belief Assignment 3/3

- Coarsening
  - Finite frame of discernment instead of $\Delta I \subset \mathbb{R}$
  - Example: $\Omega = \{wet, cloud, sun\}$
  - Mass function:

$\Delta I << 0$	$\Delta I < 0$	$\Delta I \approx 0$	$\Delta I > 0$	$\Delta I >> 0$
				
$\{wet\}$	$\{wet, cloud\}$	$\{wet, cloud, sun\}$	$\{cloud, sun\}$	$\{sun\}$

- Combination
  - Several independent measures can be combined using the Dempster rule
- Decision
  - From mass to *pignistic* probability
    $P(A) = \sum_{\emptyset \neq B \subset \Omega} m(B) \frac{|A \cap B|}{|B|}$

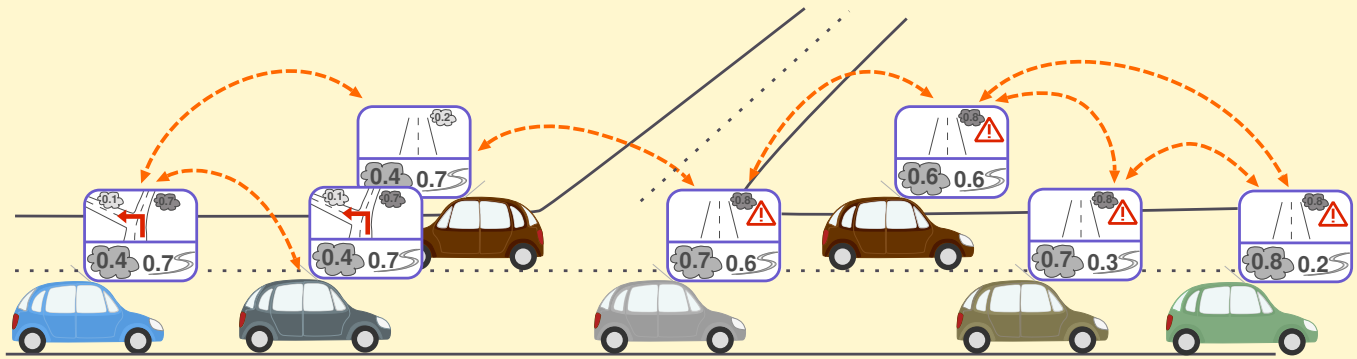
Summary

Sommaire

Distributed data fusion

Motivation 1/2

- Problem
  - Direct confidence (regularly) produced locally
    Using an external uncertain device
  - Node's confidence computed using other values
- Avoiding data collection
- Locality
  - One result per node
  - Depends on its position in the network

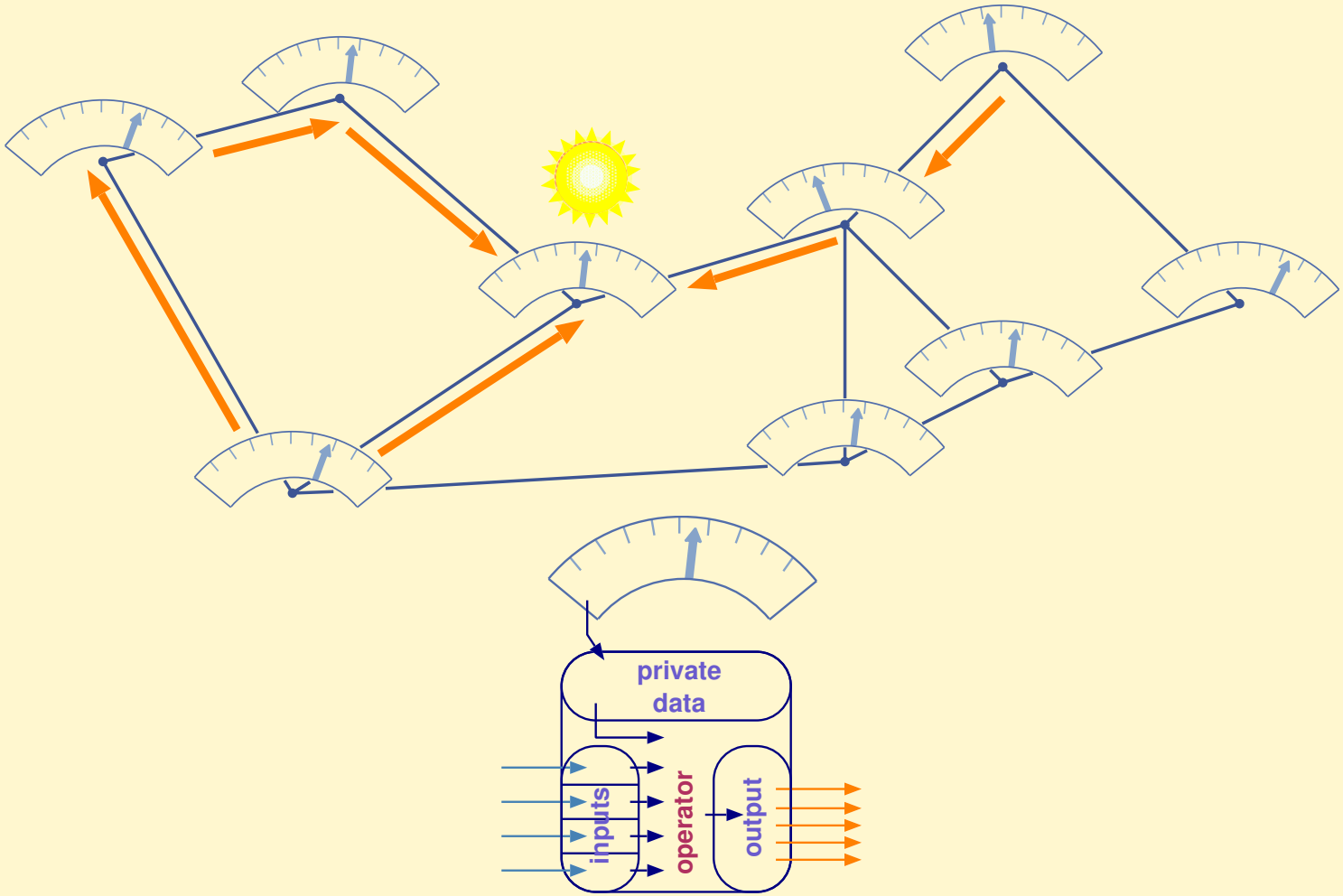


$\rightsquigarrow$ Distributed approach for data fusion

Distributed data fusion

Motivation 2/2

- Result on any node v now depends on all other nodes, not only its neighbors.



Sommaire

Distributed data fusion

Algorithm: properties

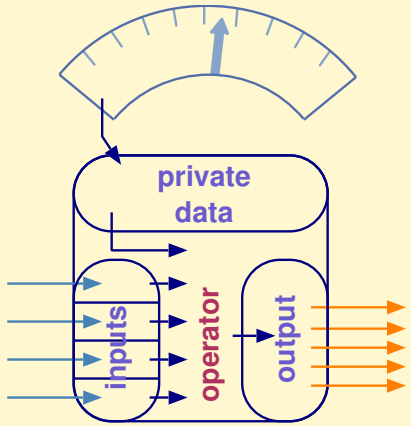
- Our distributed data fusion algorithm [SSS2012]
  - Combine all direct confidences of the system
  - Relies on local periodic broadcast
  - Discount received information
    $\rightsquigarrow$ confidence decreases according to the distance

- Properties
  - Finite data set
    Discretization + adapted operators
  - Asynchronous and anonymous system
  - Unreliable message passing system

  - Intermittent faults on memories/messages
  - Crash faults on nodes

Distributed data fusion

Algorithm: details



Upon (local) timer expiration

$\mathtt{PRIV}_v \leftarrow$ current direct confidence

$\mathtt{OUT}_v \leftarrow \mathtt{PRIV}_v$

for each entry u in $\mathtt{IN}_v$ **do**

  $\mathtt{OUT}_v \leftarrow \mathtt{OUT}_v \oplus r(\mathtt{IN}_v[u])$

end for

push($\mathtt{OUT}_v$ **)**

Reset $\mathtt{IN}_v$

Restart the timer

Distributed data fusion

Algorithm: discounting

- Local computation

  $\mathtt{OUT}_v \leftarrow \mathtt{OUT}_v \oslash r(\mathtt{IN}_v[u])$

  - $\oslash$: cautious operator defined on weights
  - r: discounting function
- Discounting **r**
  - Decreases the information
  - Application-dependent

Summary

Sommaire

Discounting

Complexity

- Local computation

  $\mathtt{OUT}_v \leftarrow \mathtt{OUT}_v \circledcirc \mathrm{r}(\mathtt{IN}_v[u])$

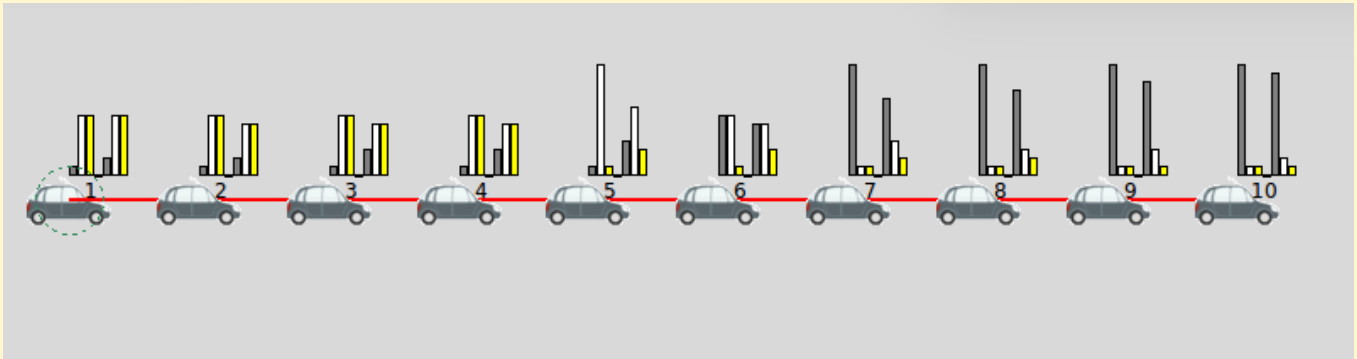
- Stabilization time — supposing a synchronous system
  - $O(k + D)$
  - k: defined by $r^k(\text{smallest value}) = \text{largest value}$
  - D: diameter of the stabilized topology

Sommaire

Discounting

Locality

- Without discounting
  - A single result per connected component
- With discounting
  - Limited influence of a node
  - Locality of the result
- Demo:



Sommaire

Discounting

Self-stabilizing proof

- r-operator [SSS2007]
  - $\triangleleft$: r-operator defined by $x \triangleleft y = x \mathbin{⍓} r(y)$
  - Condition 1: endomorphism
    $\mathbf{r}(\mathbf{w}_1 \mathbin{⍓} \mathbf{w}_2) = \mathbf{r}(\mathbf{w}_1) \mathbin{⍓} \mathbf{r}(\mathbf{w}_2)$
  - Condition 2: expansion
    $\mathbf{w} \prec_{⍓} \mathbf{r}(\mathbf{w})$

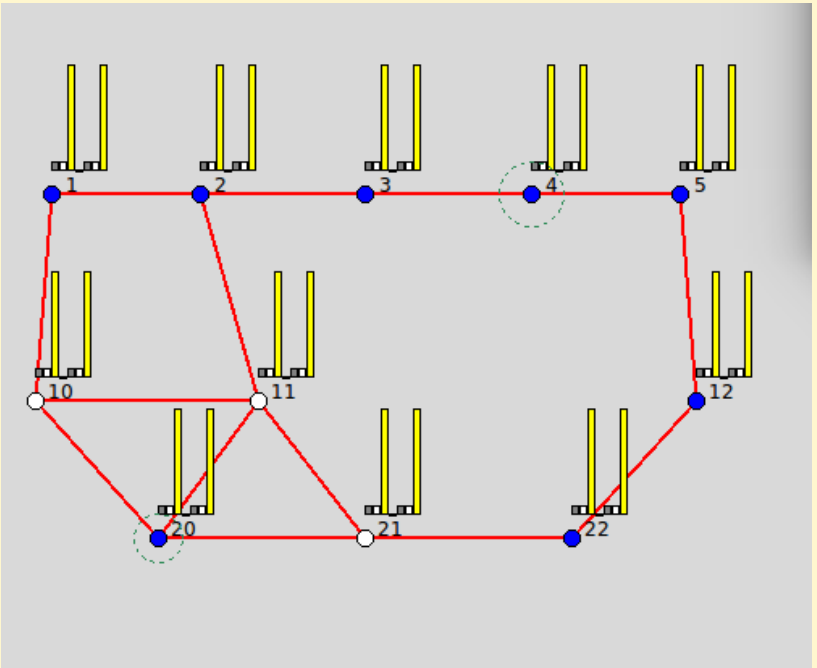
- Result [SSS 2005, SSS 2007]

  The cautious operator with the discounting r defines an r-operator which ensures the self-stabilization of the algorithm.

Discounting

Self-stabilizing proof

- Without discounting
  - No convergence after a fault
    In a message, in a memory or in the input device
- With discounting
  - Convergence in finite time after the transient failure ceases
- Demo:



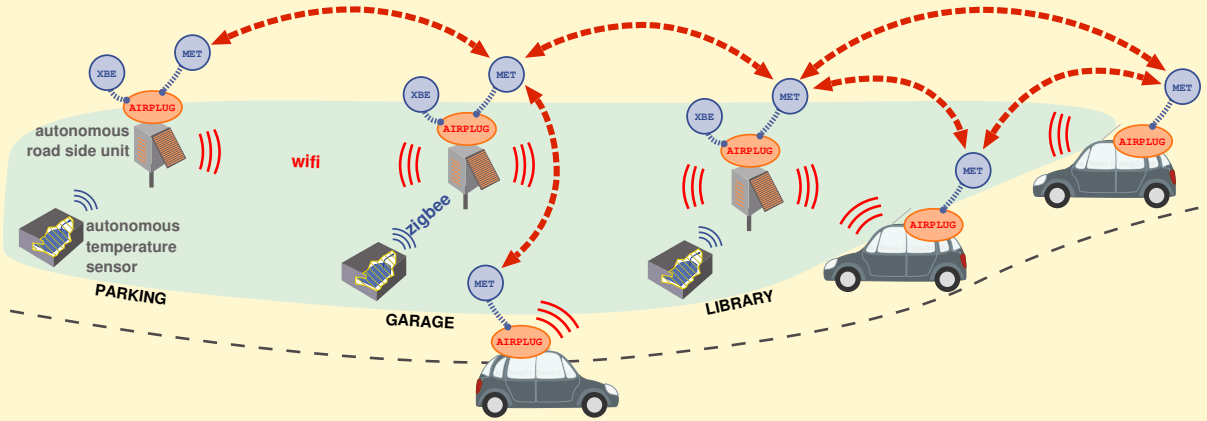
Summary

Applications

Detecting icy roads

- Testbed for icy roads detection [WiSARN 2014]
  - 3 RSU, 6 sensors + vehicles

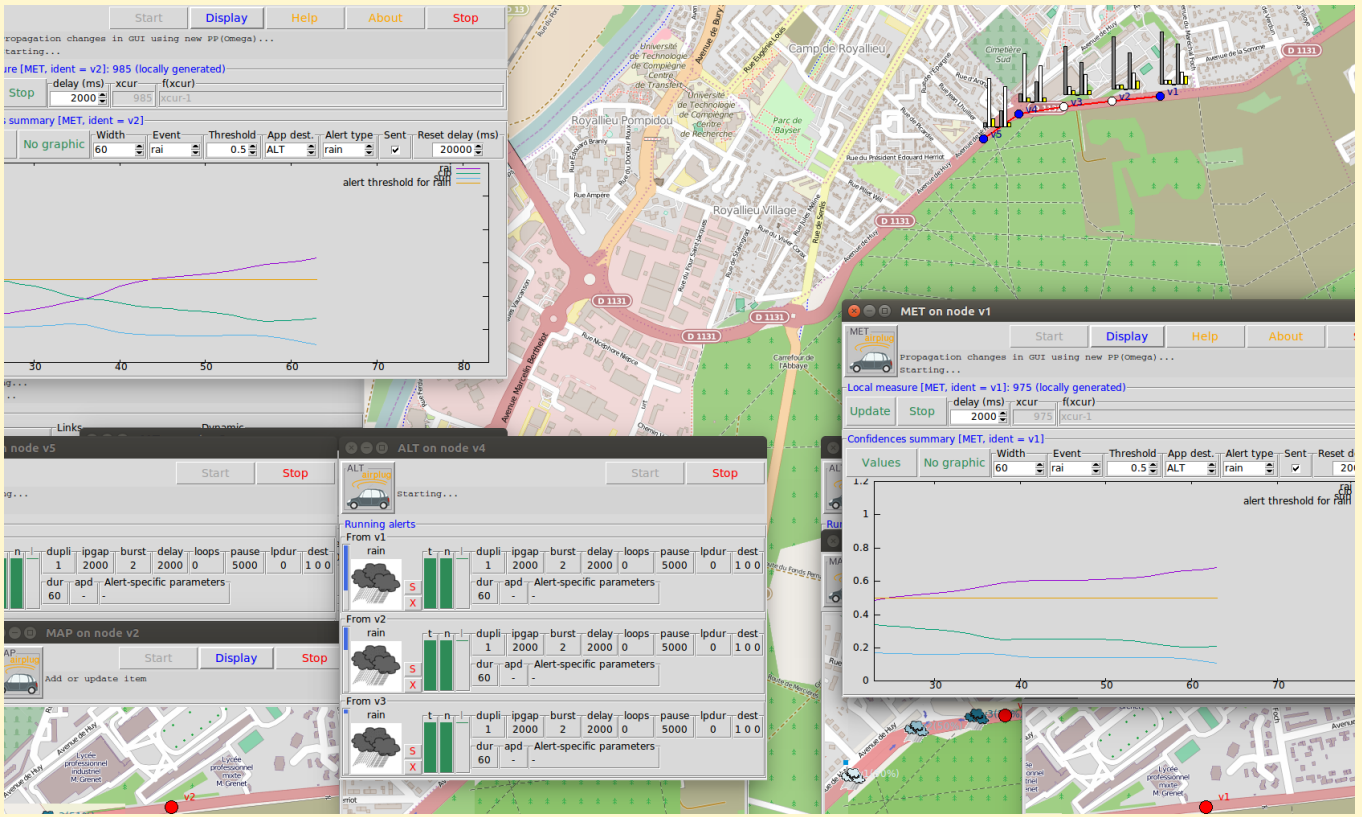
- Proof of concept [VNC 2014, IEEE ITS2016]



Applications

Detecting heavy rains on the roads

- Distributed data fusion
  - Enforce confidences in the rain event
- Decision phase [ITSC2015]
  - Pignistic probability
  - Exceeding the threshold $\rightsquigarrow$ generate an alert



Summary

Experiments with Self-Stabilizing Distributed Data Fusion

- Data fusion
  - Dealing with imprecise and uncertain data
  - Belief Function framework [Dempster Shafer]
- Distributed data fusion
  - Avoiding the data collection phase
  - One result per node depending on its position
  - Self-stabilizing algorithm [SSS 2012]
    - Cautious operator $\rightsquigarrow$ saves the data incest
    - Discounting $\rightsquigarrow$ decreases the information
- Discounting:
  complexity, locality, self-stabilization
- Future work
  - Designing new applications
  - Studying other operators