Similarity detection for smart and transparent long-range IoT relaying

ISCC 2019
Hotel Catalonia Barcelona Plaza, Barcelona, Spain
June 30th – July 3rd, 2019

Authors: Congduc Pham, Abdallah Makhoul and Mamour Diop

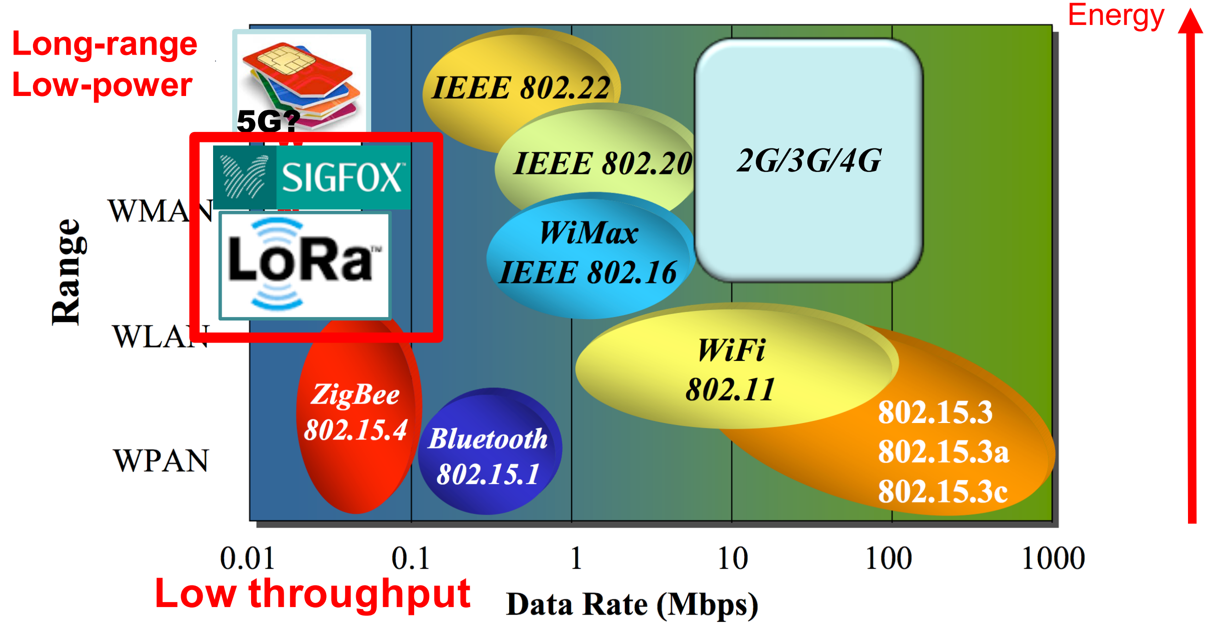
Presented on July 2nd, 2019 by C. Pham

Prof. Congduc Pham
http://www.univ-pau.fr/~cpham
Université de Pau, France

IoT – from idea to reality

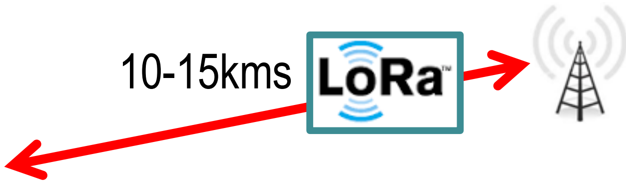
LoRa LPWAN wireless technology

Energy-Range dilemma



Semtech's LoRa provides low-power long-range transmission enabling several years of operation on batteries

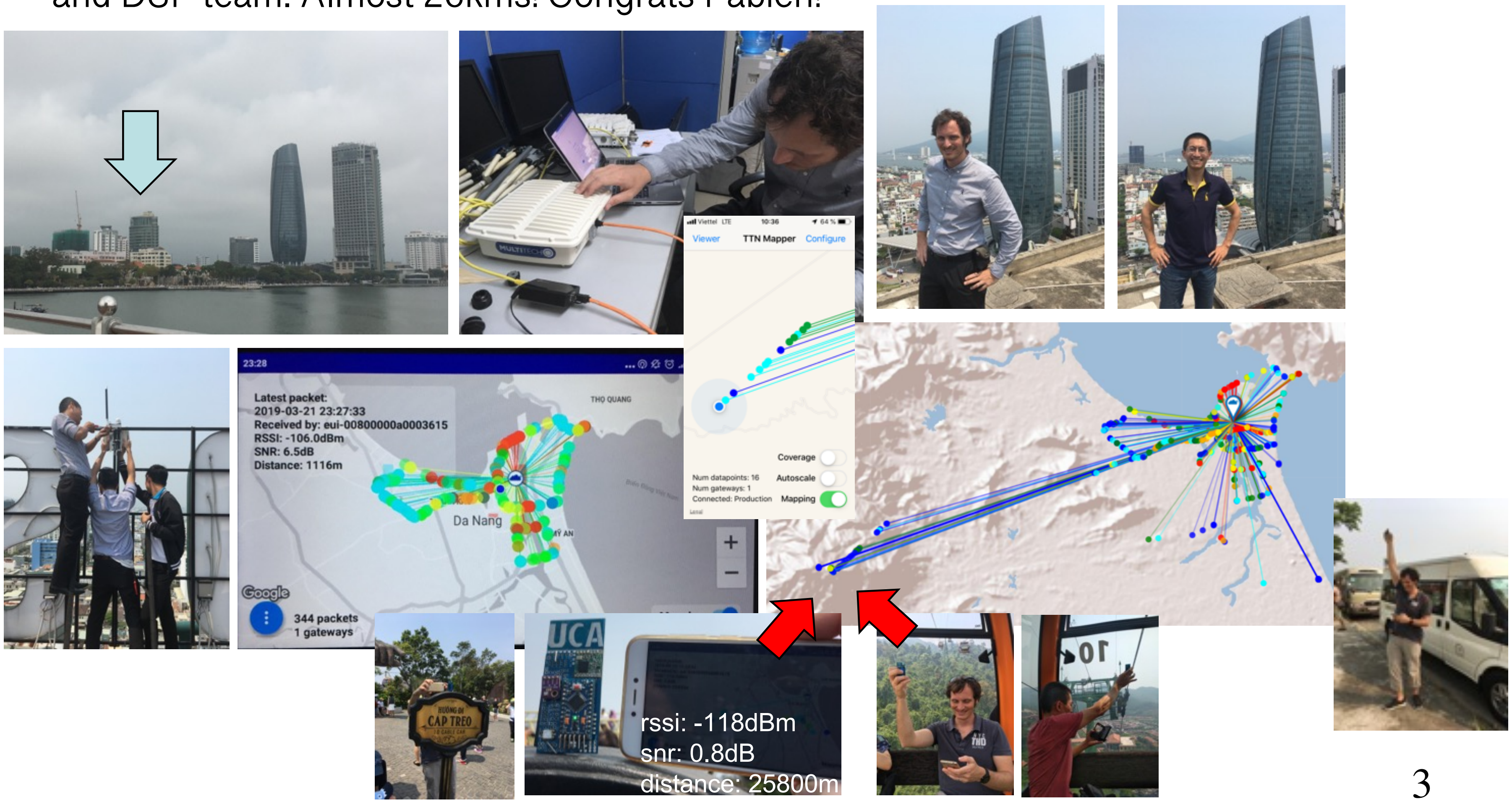
Soil moisture monitoring



LoRa networks are 1-hop, gateway-centric with possible roaming

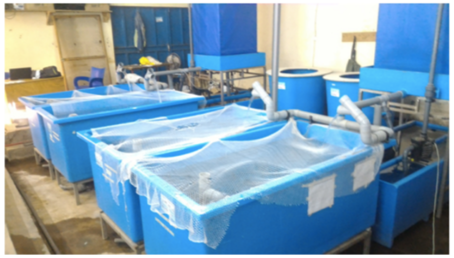
High building=large coverage

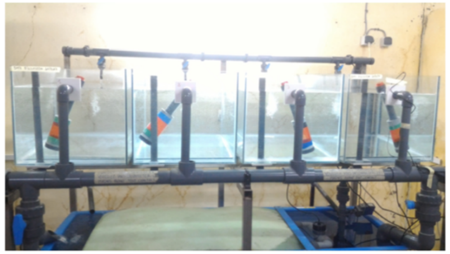
- LoRaWAN gateway on top of Danang's DSP building by Fabien, U. Danang and DSP team. Almost 26kms! Congrats Fabien!



HATCHERY EXPERIMENT, BURKINA FASO

- ❑ Laboratory named Laboratoire d'Études des Ressources Naturelles et des Sciences de l'Environnement (LERNSE)
- ❑ NAZI BONI University in a small village of Bobo-Dioulasso city
- ❑ Sensors are placed in a hatchery and the box is placed outside of the building





LOW-COST BUOY FOR FISH FARMING

In Sub-Saharian Africa, the volume of natural captured fish doesn't meet half of the population demand

Increasing production of aquaculture will help reduce the quantity of imported fishes in Africa

The aim is to monitor in real-time different parameters to control water quality and prevent some diseases that could affect fish in order to improve the quality and quantity of the production

KUMAH FARM, GHANA

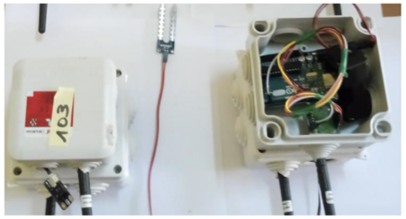
- ❑ The Kwame Nkrumah University of Science and Technology (KNUST)
- ❑ Located on the campus of the Kwame Nkrumah University of Science and Technology in Kumasi, Ghana.
- ❑ The farm comprises 30 constructed fish ponds, a farm house, a recirculating aquaculture system (RAS) laboratory and store houses.

SANAR FARM, SENEGAL

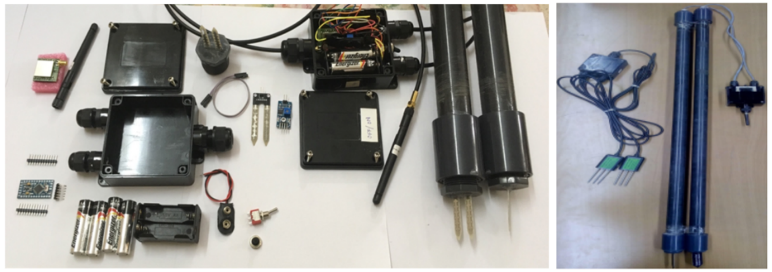
- ❑ Farm located at less than 2 km from UGB.
- ❑ One pond is dedicated for the Waziup application : 50x25m, average depth of 0.5 meters, populated by 4000 individuals of saltwater tilapia.
- ❑ The basin is irrigated via a water supply system fed by a river in proximity.
- ❑ The water in the pond is changed every 10 days

UBG farm, Senegal

Soil humidity sensor for agriculture

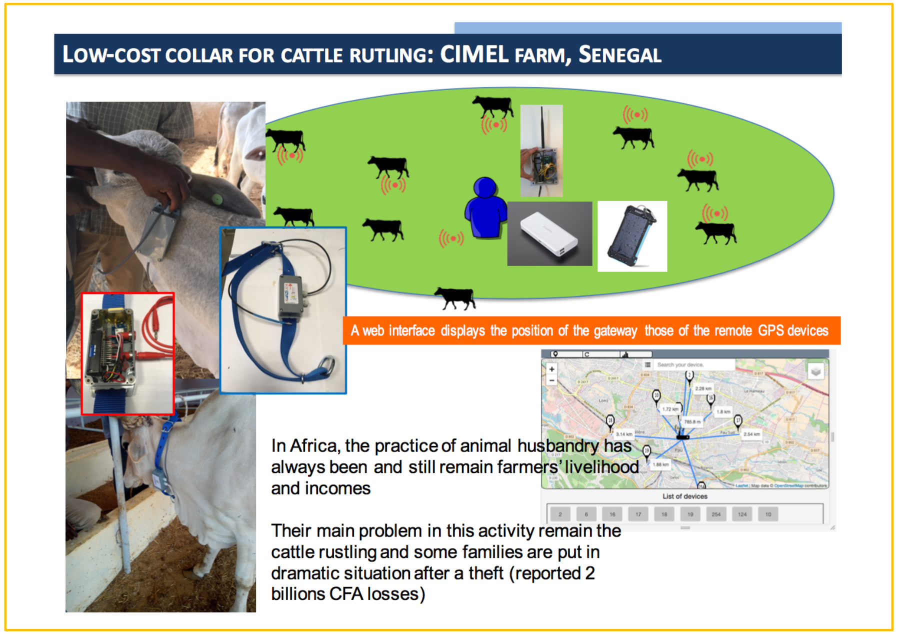


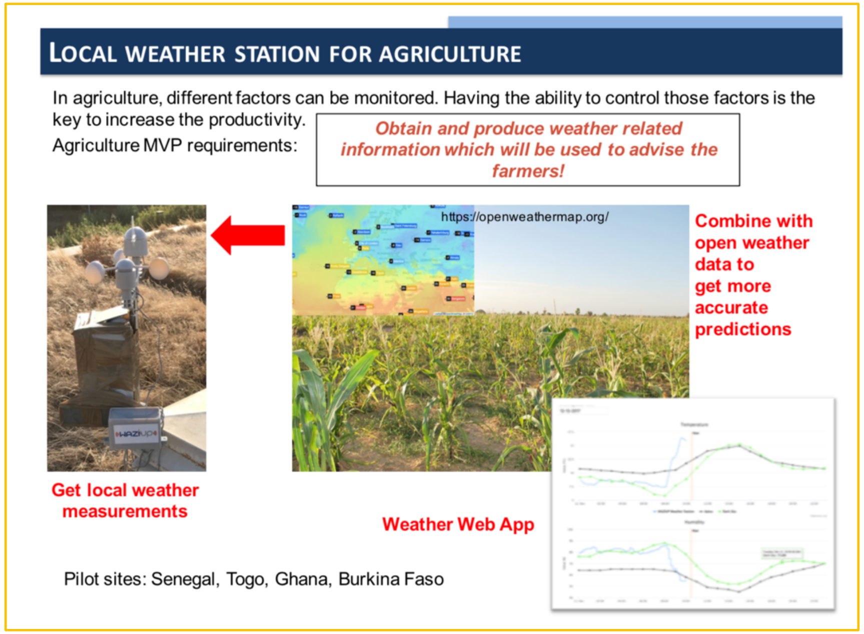
Monitoring soil moisture and other parameters to provide insightful recommendations and notifications to farmers, and advisors

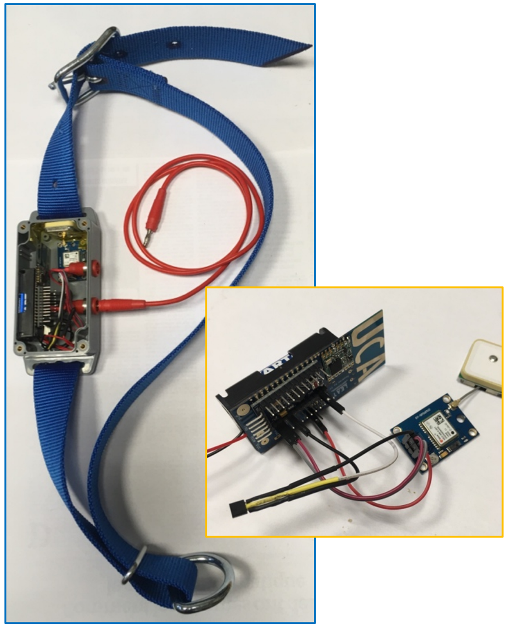


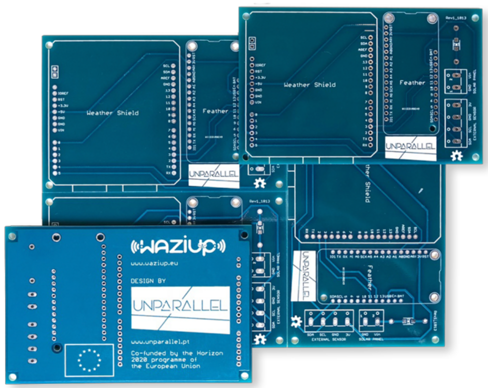
Nasso site, Burkina Faso

Urbannatic gardens, Togo









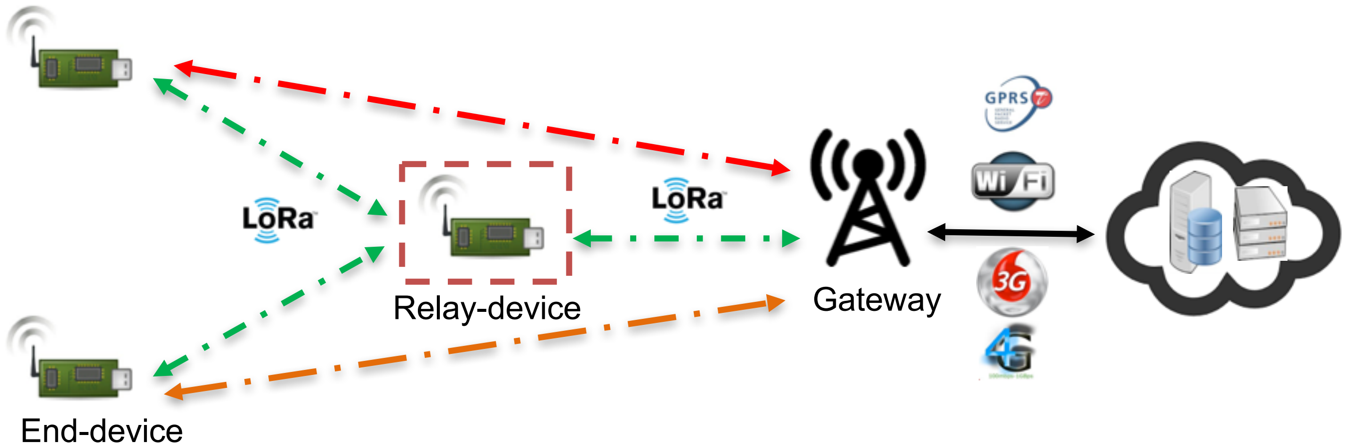
From Unparallel for WAZIUP

Deployment in rural areas no high building ☹

- Expected range: about 2-4kms
- 1-hop connectivity to gateway is difficult to achieve in real-world, remote, rural scenarios

2-hop long-range approach

- **smart, transparent** relay node should be able to be inserted at anytime between end-devices and gateway to increase range



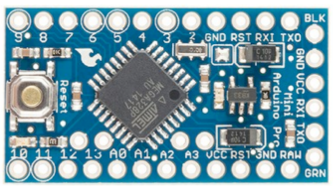
- **2 possible approaches**
  - Use periodic & short Channel Activity Detection (CAD) to detect uplink messages (recent draft from Semtech)
  - Use an observation phase (full receive mode) to determine device's schedule

Prof. Congduc Pham
http://www.univ-pau.fr/~cpham

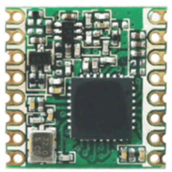
Our relay's design choices

- Observation phase + data forwarding phase
  - CAD reliability decreases as distance increases
  - A CAD returning false does not mean that there is no activity!
- On-the-fly learning of incoming traffic from end-devices: **observation phase**
- Just-in-time wake up in **data forwarding phase**
- Deep sleep between 2 wake up
- No additional hardware ➔ low-cost sensor nodes can be recycled as relay node

Arduino Pro Mini @3.3V



RFM95W

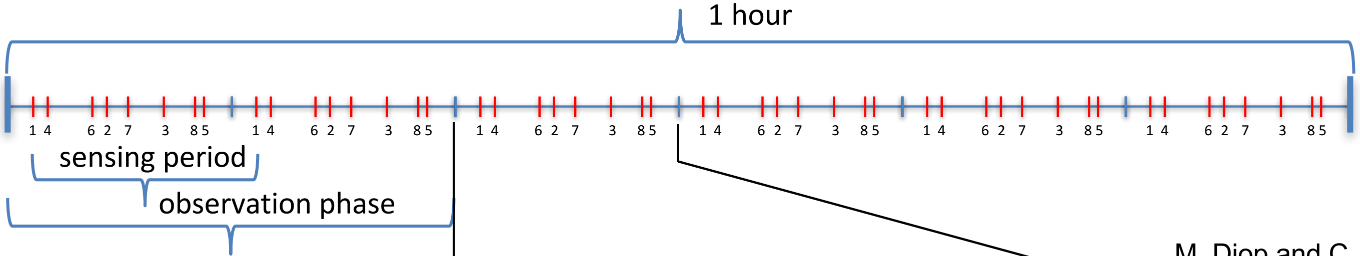


Observation phase

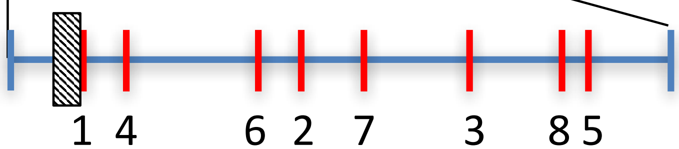
- Device i wakes up and transmit every I_target_i
  - Target TX time for device i: T0_i+n*I_target_i
  - Real TX time for device i: T0_i+n*I_real_i
- I_real_i from device i is determined during observation phase

Device i	I_real_i	ToA_i
Device j	I_real_j	ToA_j
Device k	I_real_k	ToA_k

Note that I_real_i can also take into account pkt collisions that are resolved with some kind of back-off procedure



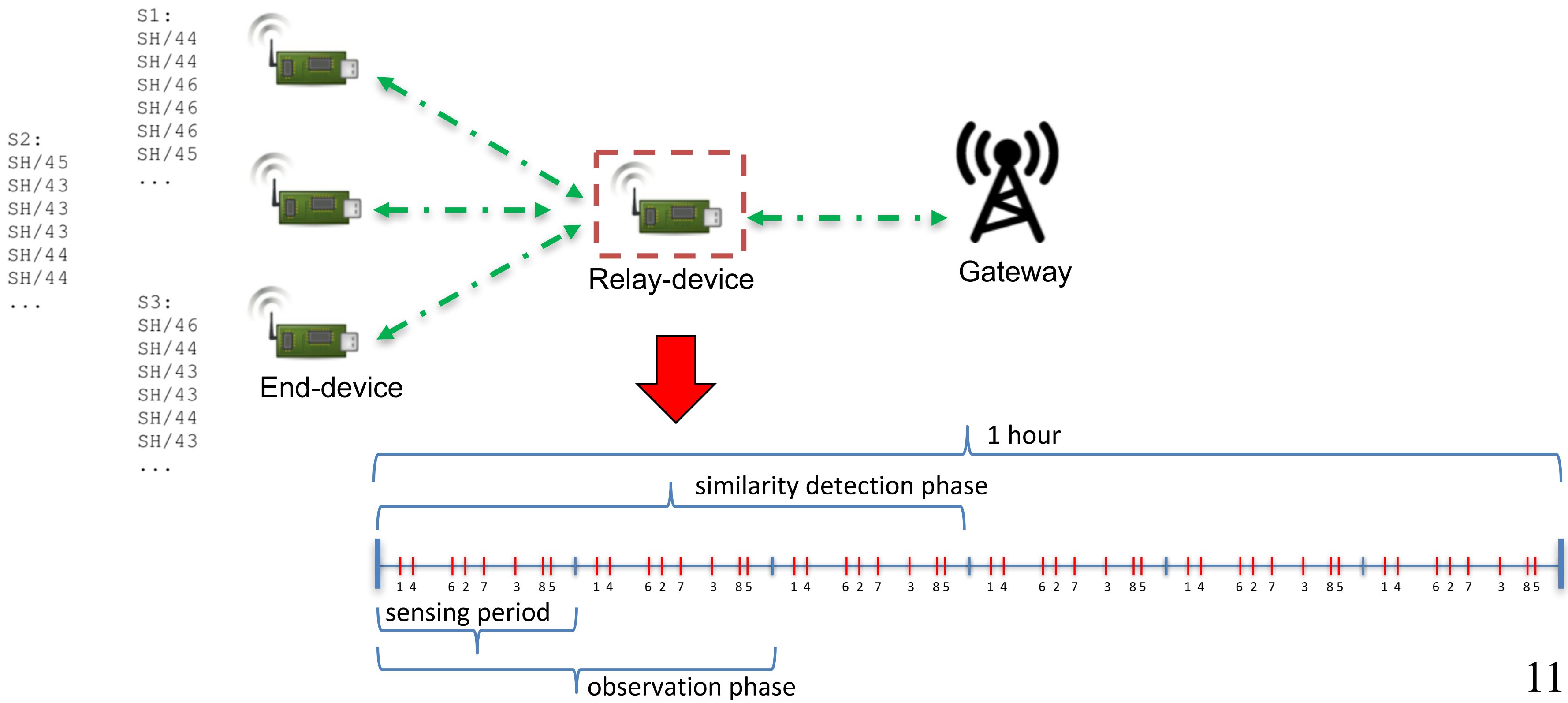
- Minimize guard time



M. Diop and C. Pham, "Increased flexibility in long-range IoT deployments with transparent and light-weight 2-hop LoRa approach", 11th Wireless Days conference, Manchester, UK, April 23-25, 2019.

Extending with similarity detection

- Find similarities between measures to avoid both waking-up and transmission to gateway
- Reduce energy consumption + help enforcing duty-cycle



Similarity function

- Many functions can be used to compute a similarity level
- Example with Euclidian distance

$$R_i = [r_{i_1}, r_{i_2}, r_{i_3}, r_{i_4}, \ldots, r_{i_\mathcal{T}}]$$
$$R_j = [r_{j_1}, r_{j_2}, r_{j_3}, r_{j_4}, \ldots, r_{j_\mathcal{T}}]$$

$$E_d(R_i, R_j) = \sqrt{\sum_{k=1}^{\mathcal{T}} (r_{i_k} - r_{j_k})^2}$$

$$\mathbb{E}_d = \{E_d(R_1, R_2), E_d(R_1, R_3), \ldots, E_d(R_{N-1}, R_N)\}$$

- Gaussian normalization

$$E'_d(R_i, R_j) = \frac{E_d(R_i, R_j) - \overline{Y}}{6 \times \sigma} + \frac{1}{2}$$

$$\overline{Y} = \frac{\sum_{k=1}^{|d|} d_k}{|d|}$$ Mean of all distances

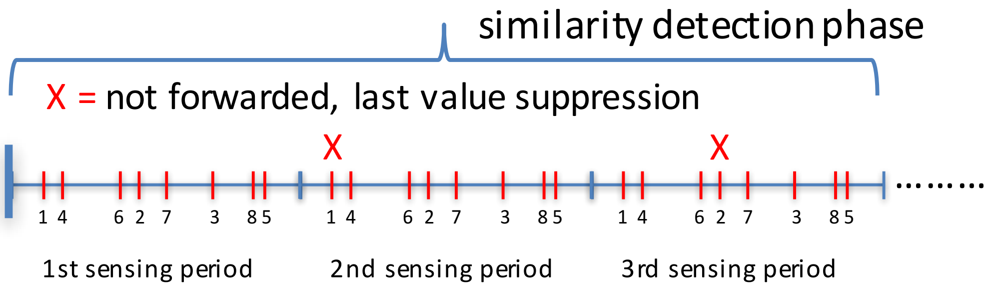
$$\sigma = \sqrt{\frac{\sum_{k=1}^{|d|} (d_i - \overline{Y})^2}{|d|}}$$ Standard deviation

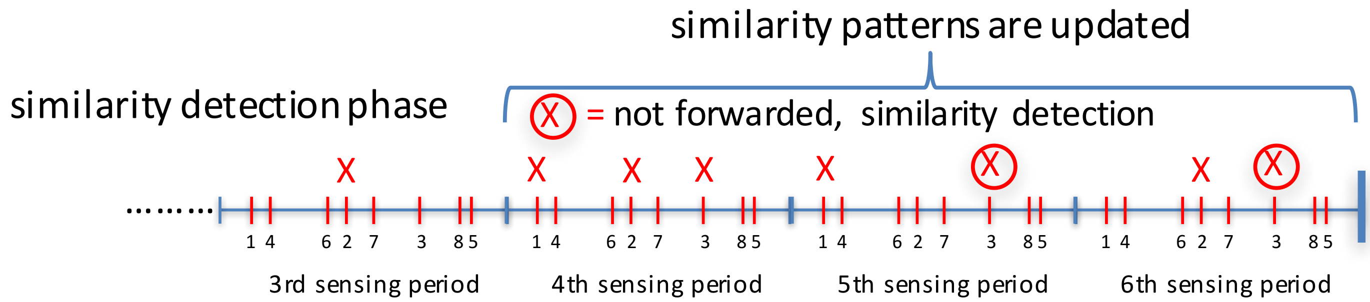
$$|d| = \frac{n \times (n-1)}{2}$$

$$E'_d(R_i, R_j) \leq \epsilon, \text{ where } \epsilon \ in[0, 1]$$

Avoiding message forwarding

```
S1:      S2:
SH/44    SH/45
SH/44X   SH/43
SH/46    SH/43X
SH/46    SH/43
SH/46    SH/44
SH/45    SH/44
...      ...
```





S1	S2	S3
44	45	46
44X	43	44
46	43X	43
46	43	43
46	44	44
45	44	43

$E'_d(R_2,R_3)=0.266 < 0.3$

S1	S2	S3
44	45	46
44X	43	44
46	43X	43
46X	43X	43X
46	44	44
45	44	43

$E'_d(R_2,R_3)=0.265 < 0.3$

S1	S2	S3
44	45	46
44X	43	44
46	43X	43
46X	43X	43X
46X	44	(44)
45	44	43

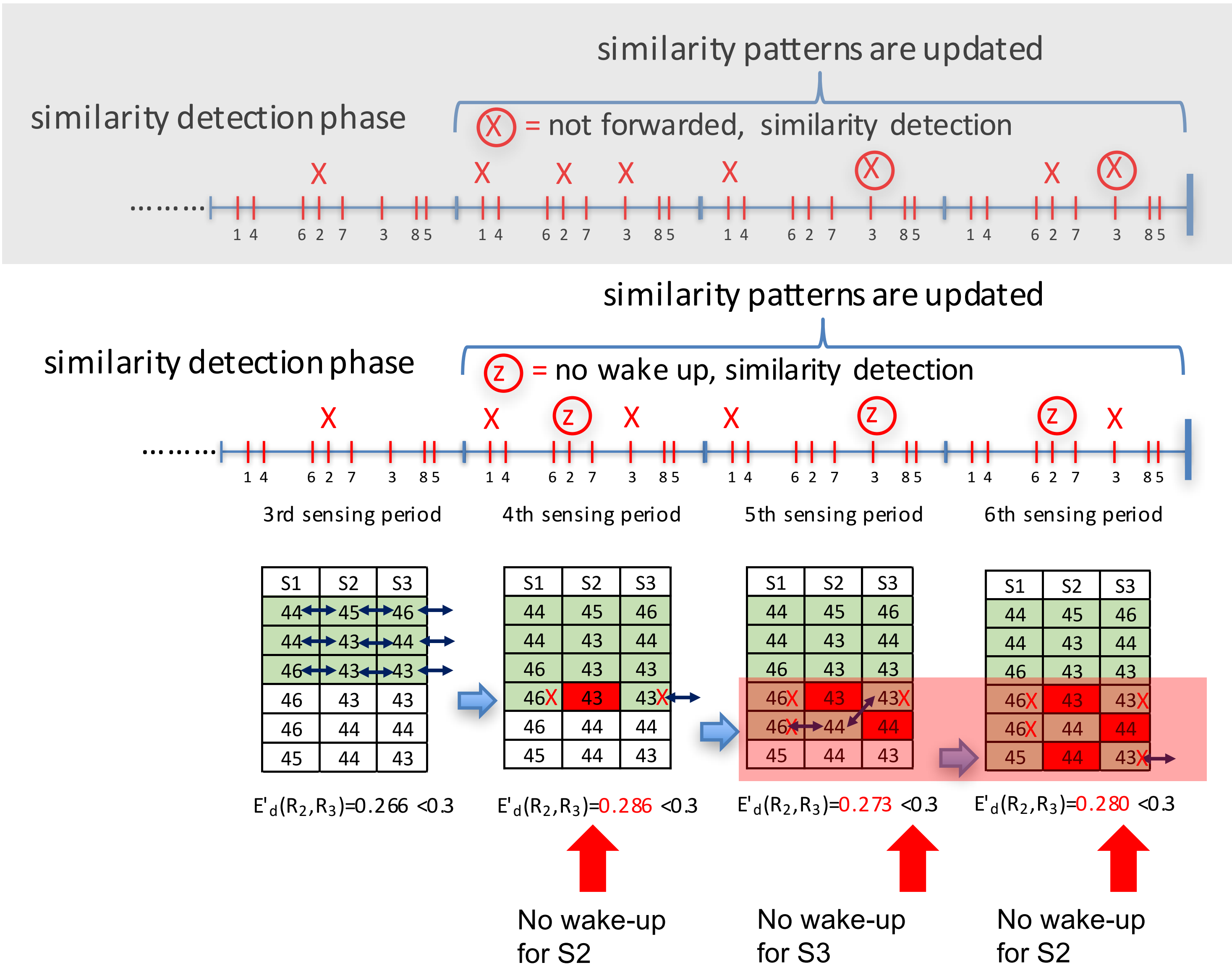
$E'_d(R_2,R_3)=0.264 < 0.3$

S1	S2	S3
44	45	46
44X	43	44
46	43X	43
46X	43X	43X
46X	44	(44)
45	44X	(43)

$E'_d(R_2,R_3)=0.266 < 0.3$

After the initial similarity detection phase (e.g. with $m = 3$), some redundancy patterns can already be identified and the relay can decide to not forward some messages. With the previous datasets example, the relay computes $\mathbb{E}_d = \{3.31, 3.60, 1.41\}$ with Eq. 1 and computes $E'_d(R_1, R_2) = 0.59$, $E'_d(R_1, R_3) = 0.64$ and $E'_d(R_2, R_3) = 0.26$ with Eq. 2 where $\overline{Y} = 2.77$, $\sigma = 0.97$ and $|d| = 3$. If the application defines $\epsilon = 0.3$ then the relay can decide that devices S2 and S3 are redundant each other.

Avoiding wake-ups



Handling missing values

- Use N/A when not waking-up for sensor i
- Replace N/A by last received value from sensor i
- Other alternatives?
- For some pairs of sensors (i,j) similarity level can be computed up to different k values

M=10

m=3

s. p.	S1	S2	S3
0			
1	44	45	46
2	44	43	44
3	46	43	43
4			
5			
6			
7			
8			
9			

s. p.	S1	S2	S3
0			
1	44	45	46
2	44	43	44
3	46	43	43
4	46	N/A	43
5			
6			
7			
8			
9			

s. p.	S1	S2	S3
0			
1	44	45	46
2	44	43	44
3	46	43	43
4	46	44	43
5	46	44	N/A
6			
7			
8			
9			

s. p.	S1	S2	S3
0			
1	44	45	46
2	44	43	44
3	46	43	43
4	46	44	43
5	46	44	43
6	45	N/A	43
7			
8			
9			

s. p.	S1	S2	S3
0			
1	44	45	46
2	44	43	44
3	46	43	43
4	46	44	43
5	46	44	43
6	45	44	43
7	45	44	N/A
8			
9			

s. p.	S1	S2	S3
0			
1	44	45	46
2	44	43	44
3	46	43	43
4	46	44	43
5	46	44	43
6	45	44	43
7	45	44	43
8	45	N/A	43
9			

Implementation issues

- How to implement the proposed approach on low-memory device (e.g. ATmega328P with only 2KB of RAM memory)
- A measurement table stores the M (e.g. M=10) last values from devices
- A similarity table of size d= n(n-1)/2 stores the similarity scores, n being the number of connected devices

	2	3	4	5	3	4	5	4	5	5
	F	F	T	F	T	F	T	F	T	F
	1				2			3		4
index	1	2	3	4	5	6	7	8	9	10

- A schedule table will be used to store the next wake-up schedule for the next periods

Schedule table

	2	3	4	5	3	4	5	4	5	5
	F	F	T	F	T	F	T	F	T	F
	1					2		3		4
index	1	2	3	4	5	6	7	8	9	10

	1	2	3	4	5	k
1	T(1)	-	-	F(1)	-	4
1=4	F(4)	-	-	T(4)	-	5

(a)

	1	2	3	4	5
2	T(1)	T(2)	F(2)	F(1)	F(2)
2=3	F(4)	F(3)	T(3)	T(4)	F(3)
2=5	-	F(5)	F(5)	-	T(5)

(b)

	1	2	3	4	5	k
3	T(1)	T(2)	F(2)	F(1)	F(2)	4
3=5	F(4)	F(3)	T(3)	T(4)	F(3)	5
	-	F(5)	F(5)	-	T(5)	6

(c)

	1	2	3	4	5
	T(1)	T(2)	F(2)	F(1)	F(2)
	F(4)	F(3)	T(3)	T(4)	F(3)
	-	F(5)	F(5)	-	T(5)

(d)

	1	2	3	4	5	k
1	T(1)	T(2)	F(2)	F(1)	F(2)	4
1=4	F(4)	F(3)	T(3)	T(4)	F(3)	5
	T(1)	F(5)	F(5)	F(1)	T(5)	6

(a)

	1	2	3	4	5
	T(1)	T(2)	F(2)	F(1)	F(2)
	F(4)	F(3)	T(3)	T(4)	F(3)
	T(1)	F(5)	F(5)	F(1)	T(5)

(b)

Conclusions

- 1-hop to GW can be challenging in real-world, rural areas
- 2-hop LoRa will provide much higher flexibility in deployment
- Embedded similarity detection mechanism
  - avoid both waking-up and transmission to gateway
  - Reduce energy consumption
  - help enforcing duty-cycle
- Euclidien distance used but other functions can be also used
- Future works
  - Include comparisons between similarity functions
  - Investigate other alternatives for handling missing values