



PRIVACY-PRESERVING USE OF INDIVIDUAL SMART METERING DATA FOR CUSTOMER SERVICES

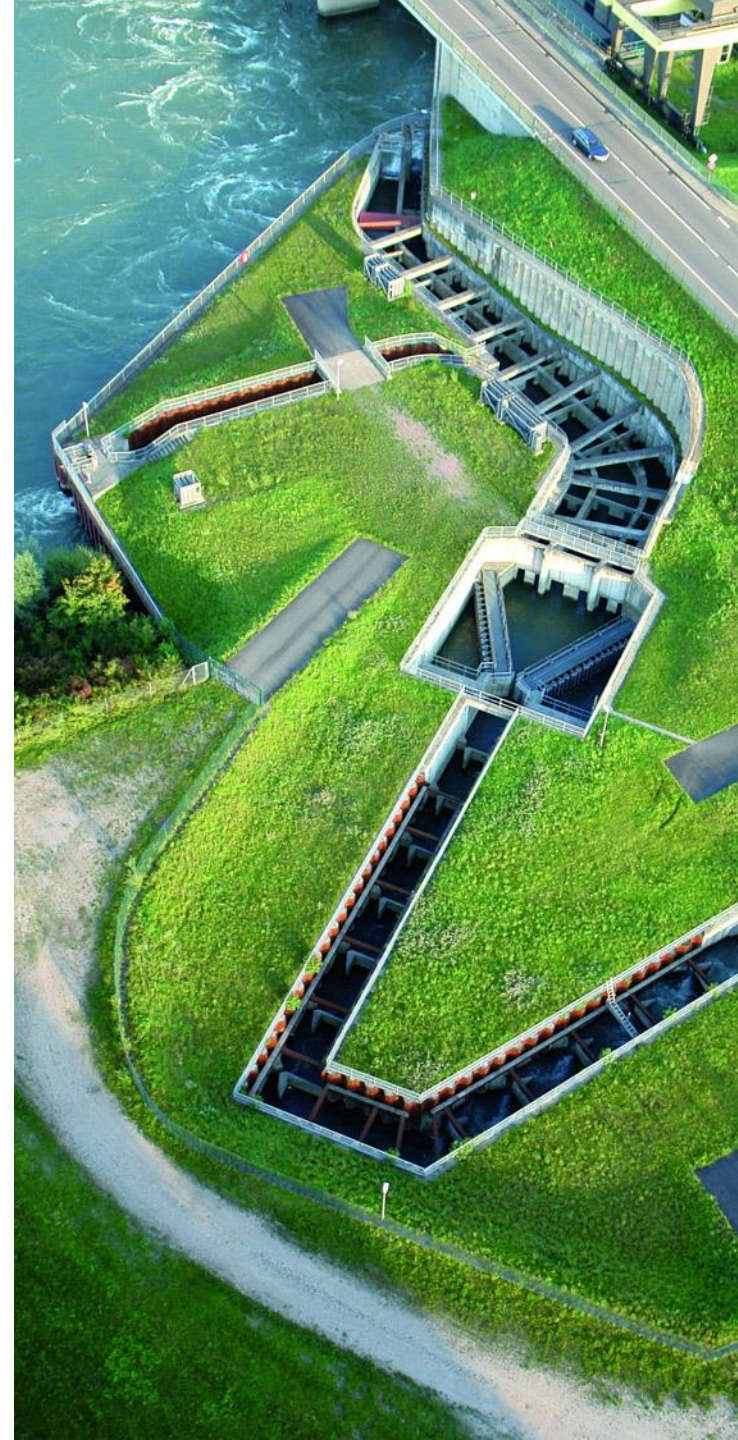
Georges Hébrail, EDF R&D

*Joint work with T.Allard, F.Masseglia, E.Pacitti
INRIA Zenith and University of Montpellier, France*

Séminaire “Data Science sur un plateau”

11 Octobre 2017

EDF Lab Saclay





SMART METERS AND CONNECTED OBJECTS

- **Deployment of smart meters** (Linky project in France)

- From 2016 to 2020 (35M meters) – 8M end of 2017
- Remote turning power on/off, remote readings and billing
- Readings up to every 10 minutes to the supplier/distributor
- Readings up to 2s on premisses



- **Deployment of connected objects in households** ('smart home')



NEW SERVICES TO CUSTOMERS

- Using smart meter readings for energy efficiency diagnosis and advice



Source particulier.edf.fr

NEW SERVICES TO CUSTOMERS

- Using smart meter readings for energy efficiency diagnosis and advice

DECOUVRIR LES EQUIPEMENTS ENERGIVORES



Source particulier.edf.fr

NEW SERVICES TO CUSTOMERS

- Using smart meter readings for energy efficiency diagnosis and advice



leaffully

It's easy to understand your energy footprint.

Leaffully is the easiest way to understand and reduce your energy footprint. Start saving today to spend less money on energy and to help the environment.

Sign up - It's free!

The image shows the Leaffully logo, which is a green leaf icon followed by the word "leaffully" in a green sans-serif font. Below the logo is the tagline "It's easy to understand your energy footprint." To the right, there is a paragraph describing the app's purpose. At the bottom, there is a green button with the text "Sign up - It's free!". On the left side of the bottom section, there is a screenshot of the app's interface on a laptop and a smartphone, showing energy usage graphs and data.

UnPlug Stuff
A Green Button App

Your Home Idles

Your home is like a car idling in the garage. While you're asleep or when you're away, devices in your home are chugging along. Even when off, they still use electricity when plugged in. What a waste.

How Much Is Your Home Wasting

The UnPlug Stuff app tells you how much energy your home is wasting when idling. As a PG&E customer, it's easy to use this app. Just [click](#) the PG&E logo to the left. Then enter your smart meter [Service Agreement ID](#) (SAID) and your online PG&E account [PIN](#). Within a few minutes you'll see your home's idle load. It's that simple.

The image shows the UnPlug Stuff app interface. At the top, there is a header with the app name "UnPlug Stuff" and the subtitle "A Green Button App". Below this, there is a section titled "Your Home Idles" with a paragraph explaining that devices in the home use electricity even when off. To the right of this section is a graphic of a house with a red roof and a green dollar sign, with the text "Your home is idling. Wasting Money". Below the "Your Home Idles" section is another section titled "How Much Is Your Home Wasting" with a paragraph explaining how to use the app. To the left of this section is the PG&E logo.

NEW SERVICES TO CUSTOMERS

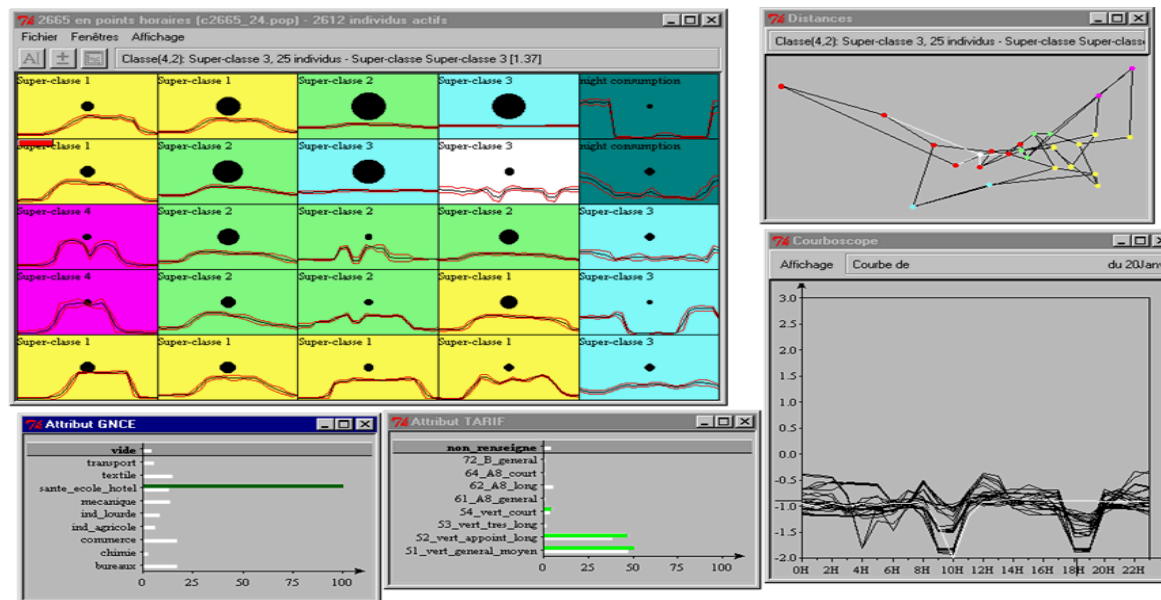
- Using smart meter readings for energy efficiency diagnosis and advice



Source www.opower.com

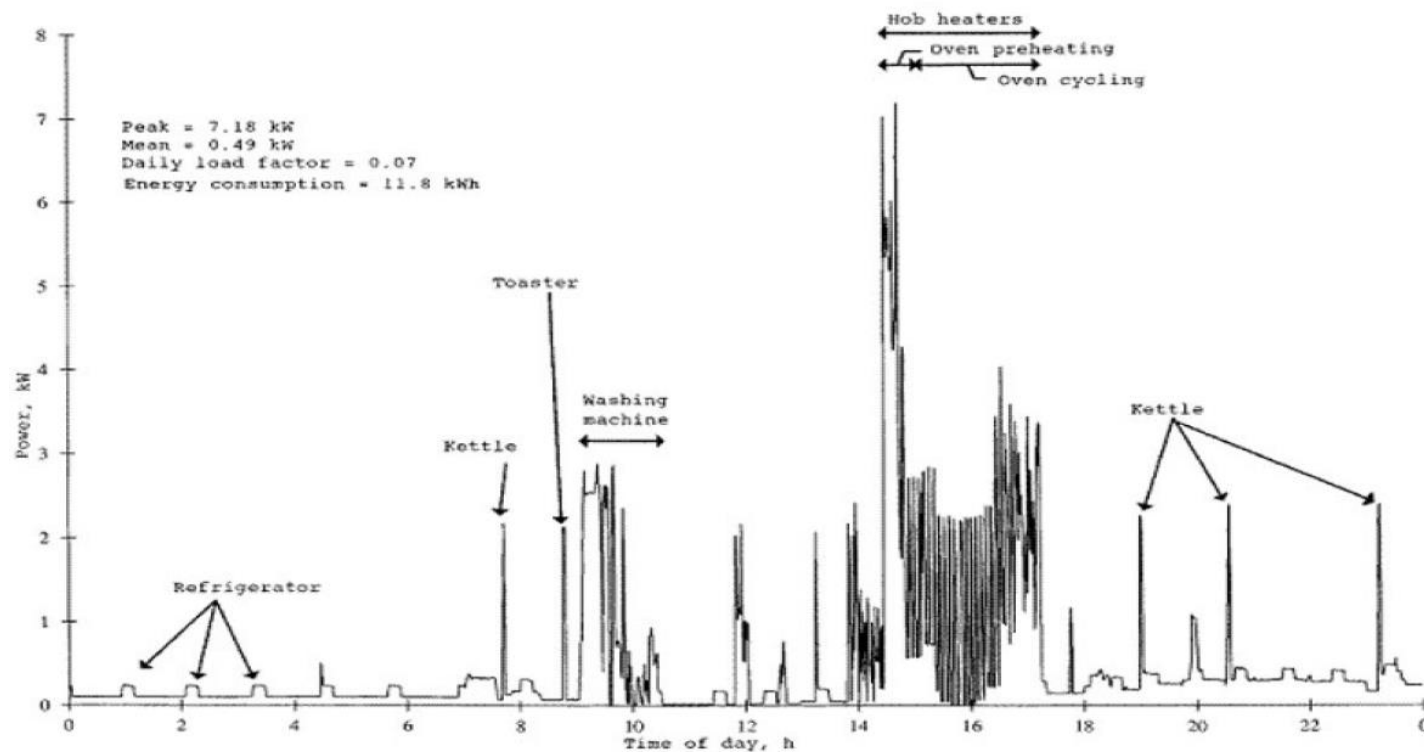
NEW SERVICES TO CUSTOMERS

- Using smart meter readings for energy efficiency diagnosis and advice
 - One standard approach: **comparison to « neighbors »**
 - Storage of individual consumption curves in a centralized data warehouse
 - Construction of (daily/weekly) profiles by clustering of individual curves
 - Association of house/equipment/occupants characteristics to clusters
 - Comparison of individual data with profiles



GREAT ... BUT ...

- Consumption data becomes more sensitive at a higher sampling rate
 - Presence/absence, number of people in the house
 - Human activity (cooking, shower, TV, ...)



Household electrical consumption example

Newborough et P. Augood, « Demand-side management opportunities for the UK domestic sector », Generation, Transmission and Distribution, IEE Proceedings-, vol. 146, n° 3, p. 283 -293, mai 1999.

PRIVACY-PRESERVING SERVICES TO CUSTOMERS

Do the same job but with privacy preservation of individual electric power consumption curve !

→ « **Chiaroscuro** »

- **Basic idea**

- Customer advice is computed locally (can easily be private)
- Construction of profiles with associated household characteristics

→ New approach of **privacy-preserving clustering of individual consumption curves**

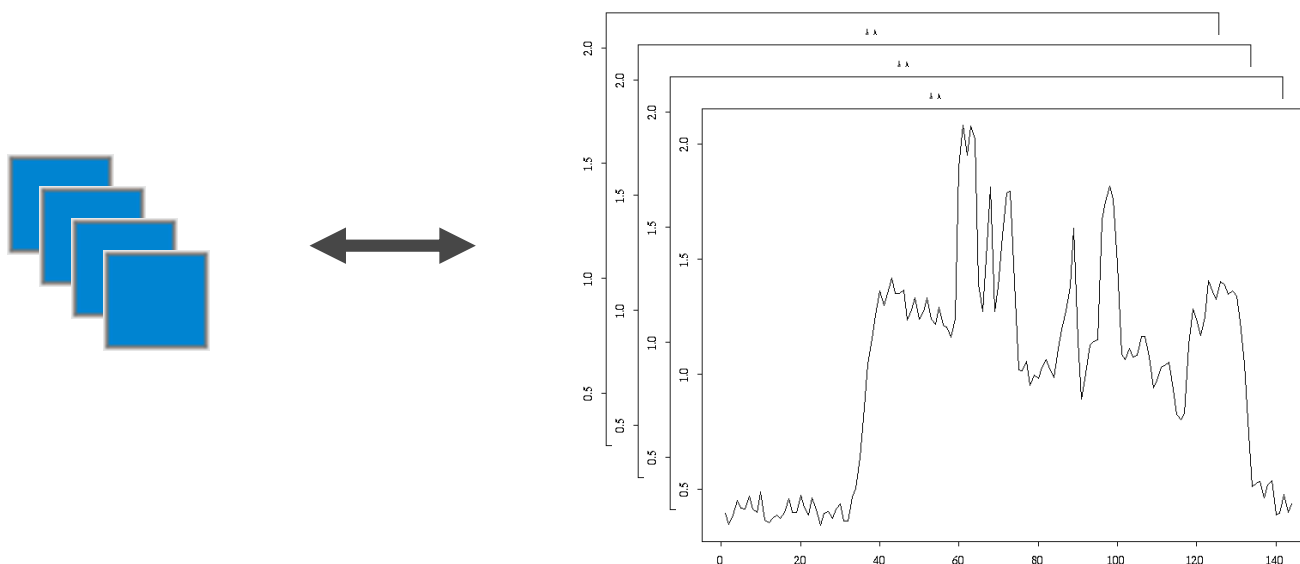
PRIVACY-PRESERVING TIME SERIES CLUSTERING

- **Privacy-preserving distributed clustering**
- **P2P infrastructure**
- **Evaluation**

PRIVACY-PRESERVING DISTRIBUTED CLUSTERING

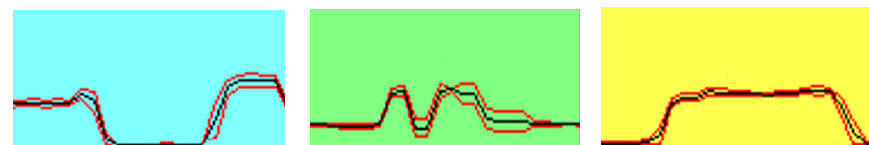
■ Data input

- N geographically distributed **individual** daily electric power consumption time series
- 24 dimensions vectors if hourly data, 144 dimensions data if 10' data
- Euclidian distance on (normalized) coordinates



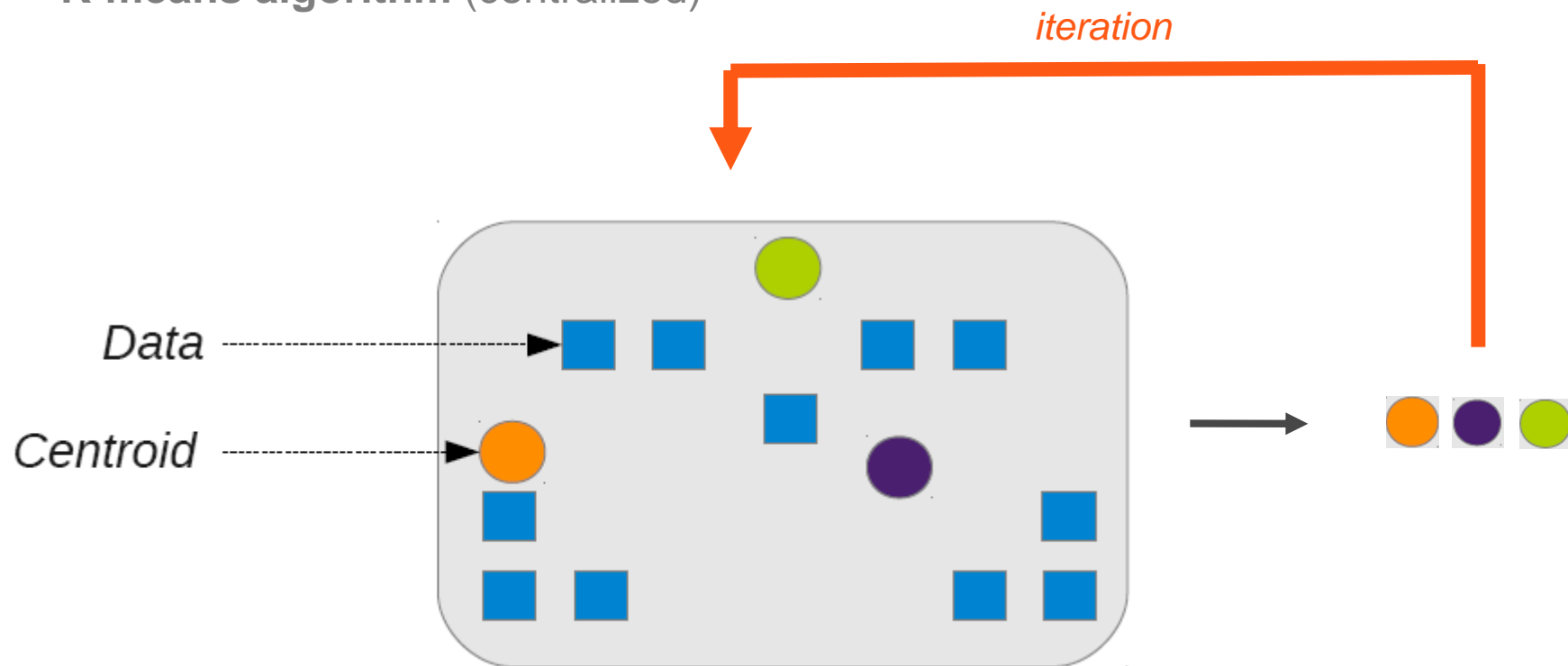
□ Output result

- K time-series **profiles** (24 ou 144 dimensions)



PRIVACY-PRESERVING DISTRIBUTED CLUSTERING

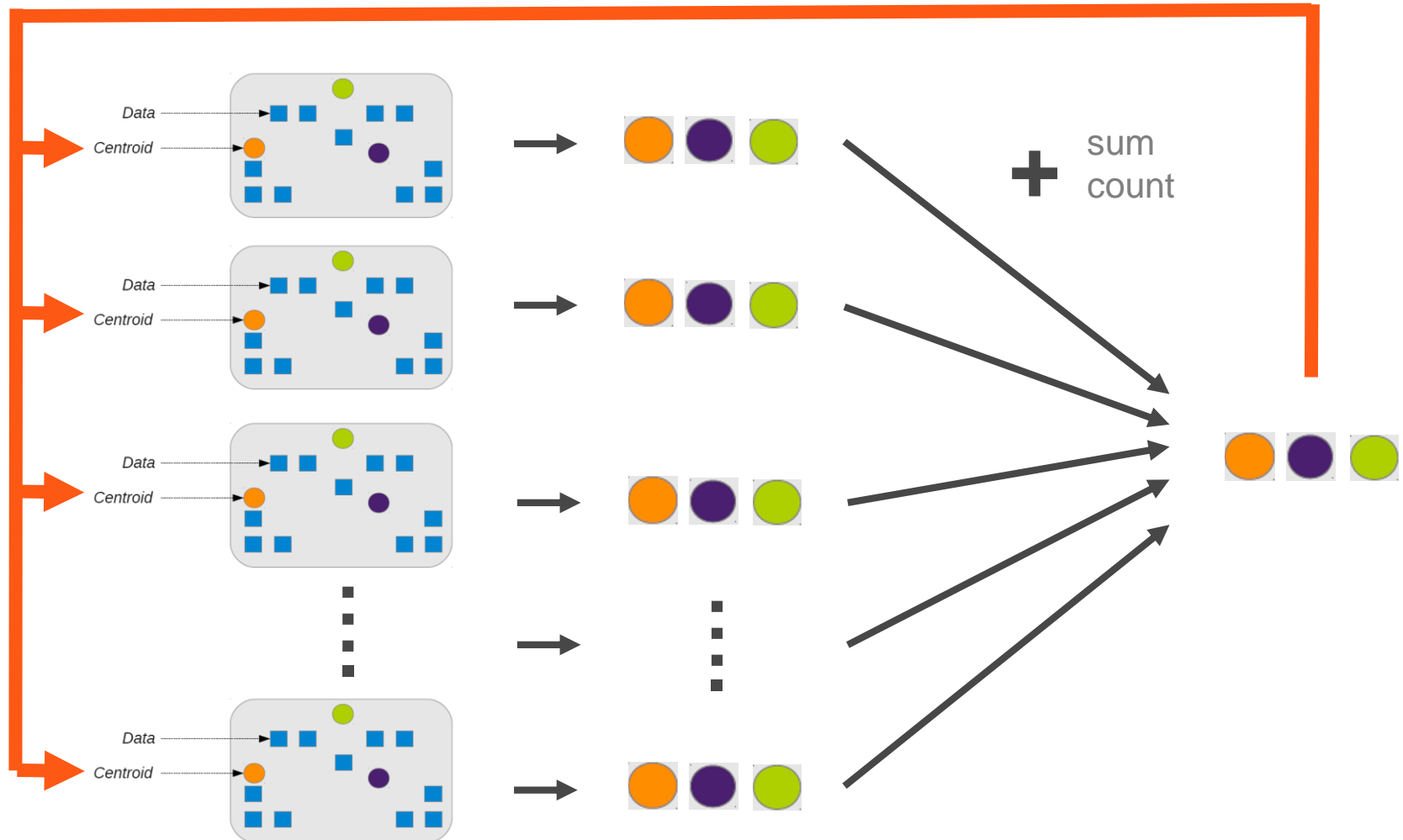
- **K-means algorithm** (centralized)



PRIVACY-PRESERVING DISTRIBUTED CLUSTERING

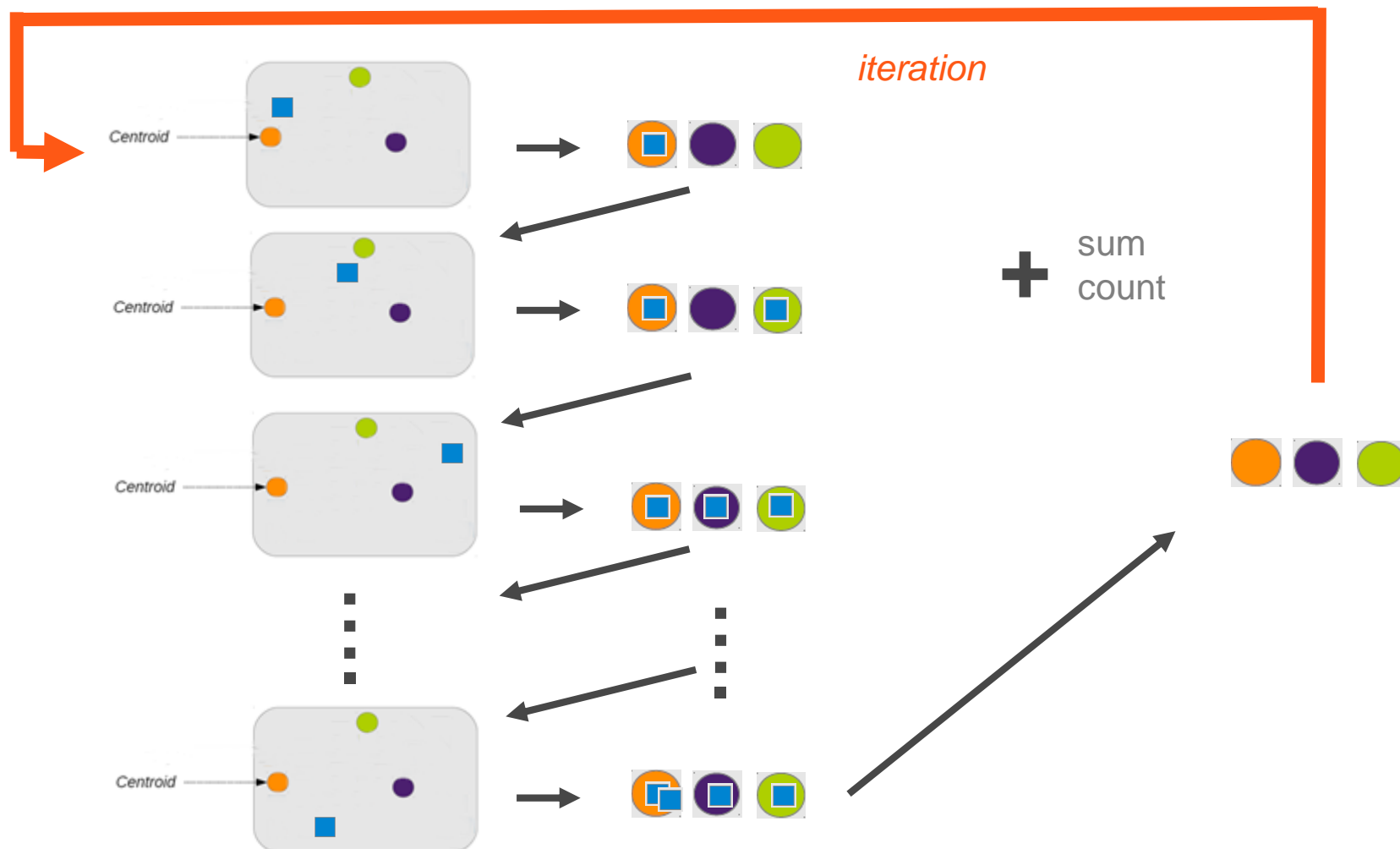
■ K-means parallelization (partition)

iteration





PRIVACY-PRESERVING DISTRIBUTED CLUSTERING

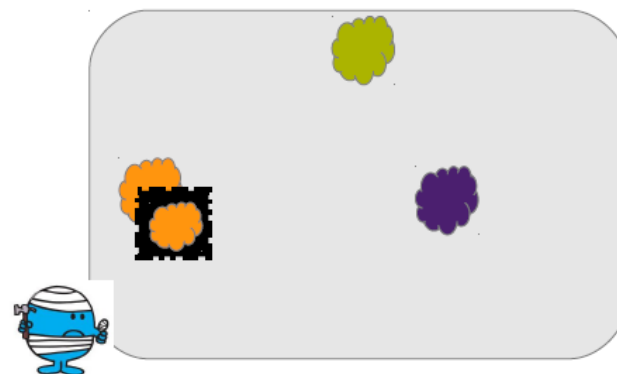
- **K-means:** *circulation* of centroids among individuals



PRIVACY-PRESERVING DISTRIBUTED CLUSTERING

- Circulation of 2 centroid structures among individual participants
 - **Cleartext** centroids for local assignment of individual time series to the closest cluster
 - **Encrypted** centroids built gradually from assignments for the next iteration

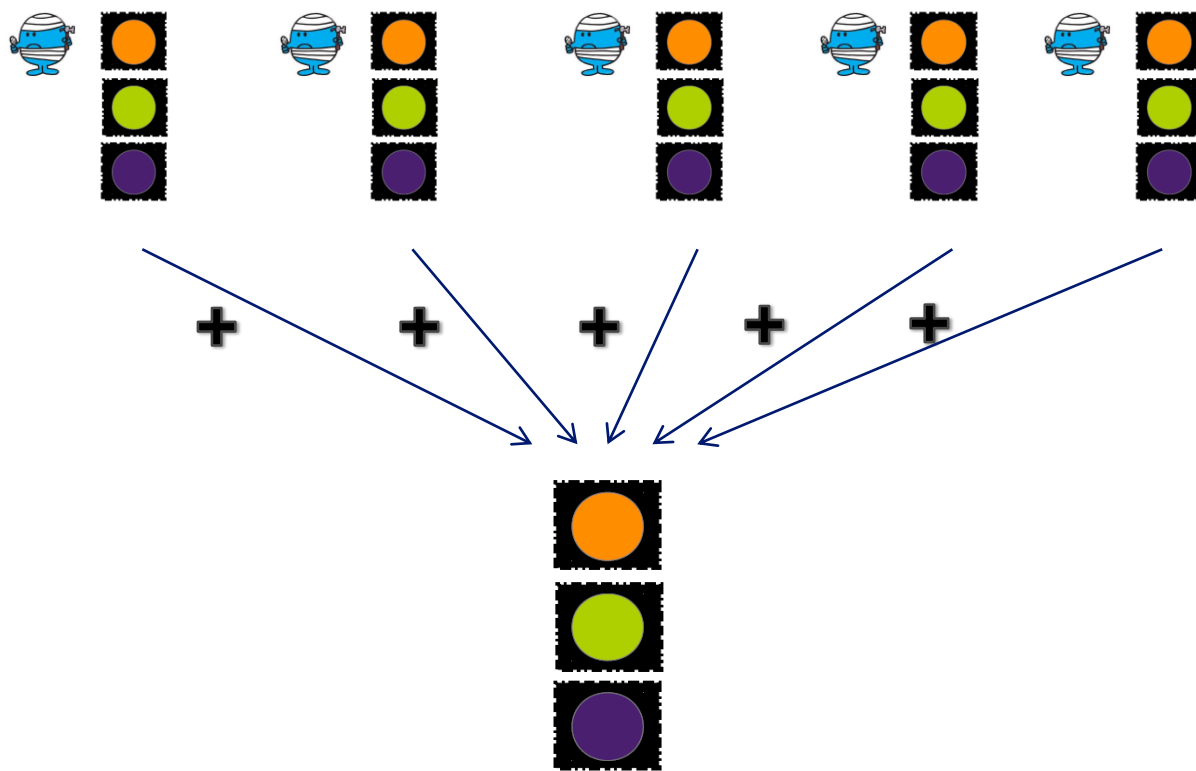
	
Cleartext centroids perturbed (differential privacy)	Encrypted means (additively-homomorphic)



PRIVACY-PRESERVING DISTRIBUTED CLUSTERING

- Centroid computation within an iteration

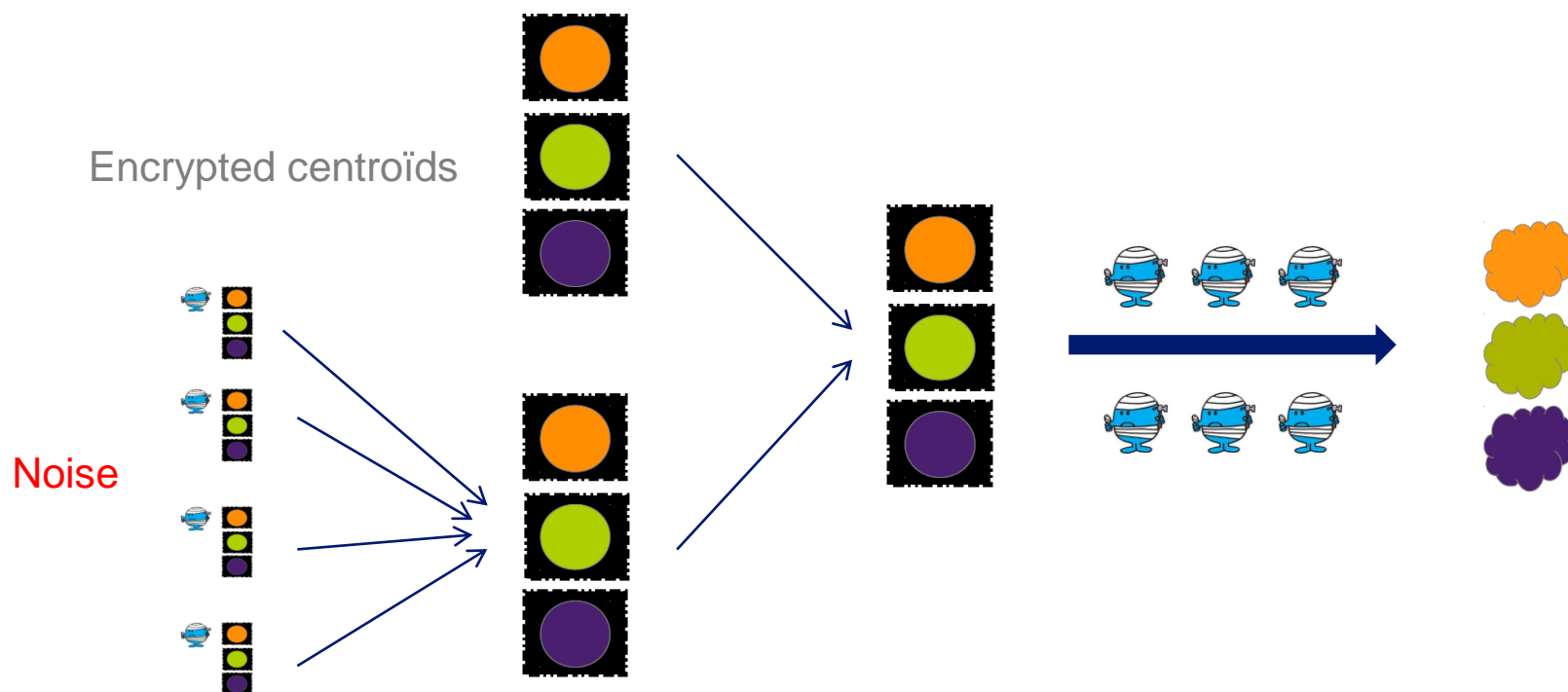
- Two additive parts: SUM and COUNT
- Use of additive **homomorphic** encryption (*allows addition directly on encrypted data*)



PRIVACY-PRESERVING DISTRIBUTED CLUSTERING

■ End of iteration

- Decryption of centroids for the next iteration but:
 - **Introduction of noise** in centroids before decryption (differential privacy)
- Collaborative decryption



PRIVACY-PRESERVING DISTRIBUTED CLUSTERING

- **Association of house/equipment/occupants characteristics to clusters**
 - Last iteration
 - Counting for each combination *characteristic x cluster*
 - Similar protection: encryption + noise + collaborative decryption

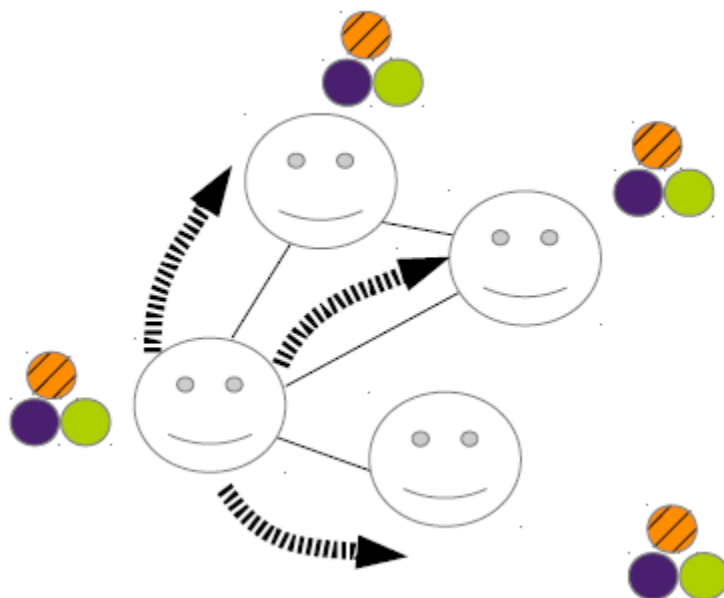
PRIVACY-PRESERVING TIME-SERIES CLUSTERING

- Privacy-preserving distributed clustering
- **P2P infrastructure**
- Evaluation

PRIVACY-PRESERVING TIME-SERIES CLUSTERING

- **P2P (peer-to-peer) architecture**

- No central server (local operations preserving privacy)
- Scalability to millions of customers
- Robustness to connections / disconnections (churn)
- Sum computations using a « **gossiping** » algorithm
 - repeated averages between participants (adaptation of usual gossip sum algorithm)



PRIVACY-PRESERVING TIME-SERIES CLUSTERING

- Privacy-preserving distributed clustering
- P2P infrastructure
- Evaluation

PRIVACY-PRESERVING TIME-SERIES CLUSTERING

- **Evaluation questions:**

- **Quality of clustering:**

- Perturbed centralized k-means implementation
 - Measured by the intra-cluster inertia
 - Datasets : Irish CER (3M real electrical consumption time-series) and NUMED (1.2M synthetic tumor growth time-series)

- **Latencies** of gossip algorithms: distributed computing simulator (Peersim)

- **Local performances** (*i.e.*, CPU times, bandwidth consumption): laptop with *current average*+ resources

PRIVACY-PRESERVING TIME-SERIES CLUSTERING

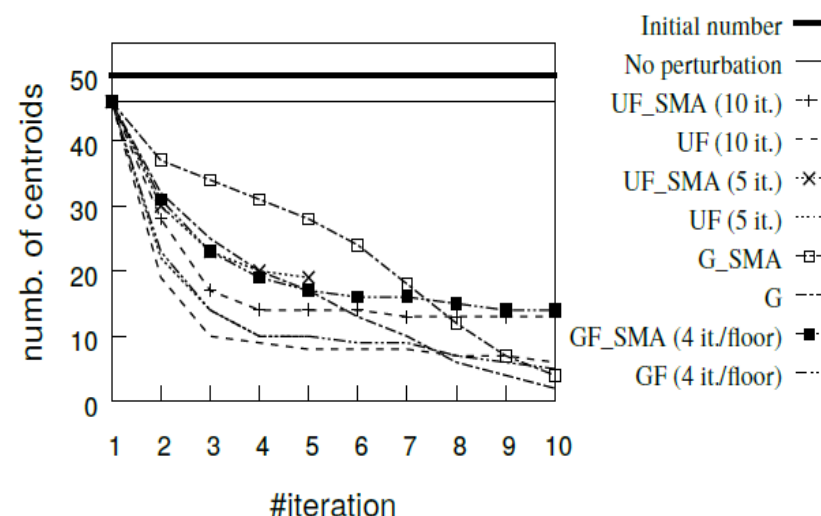
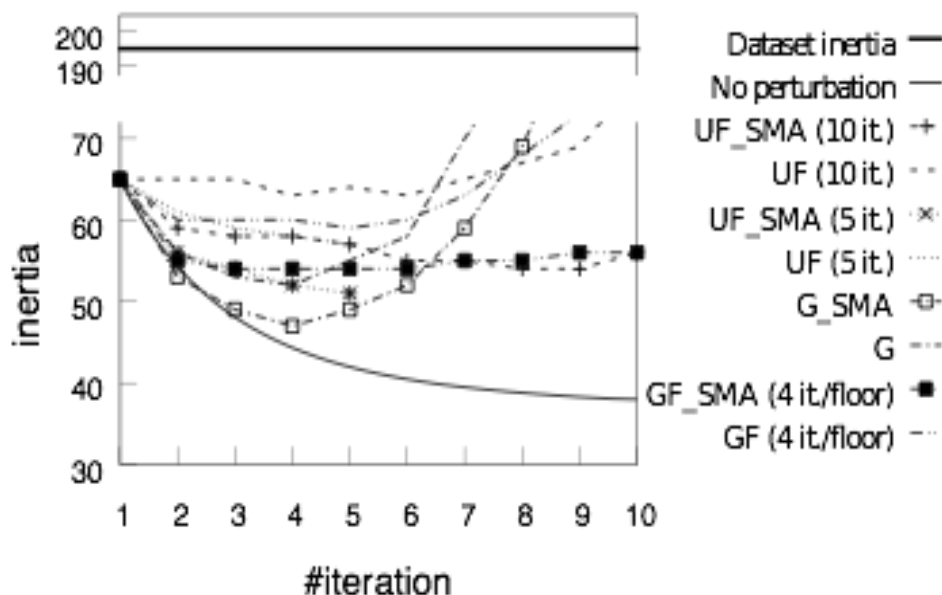
■ Quality of clustering

- Varying participants for each iteration (connections/disconnections)
- Introduction of noise
 - High perturbation for small clusters
 - Large clusters « eat » small clusters
- Distribution of privacy budget between iterations
- Smoothing time series after noise introduction
- Early stopping

PRIVACY-PRESERVING TIME-SERIES CLUSTERING

- **Quality of clustering: example of settings**

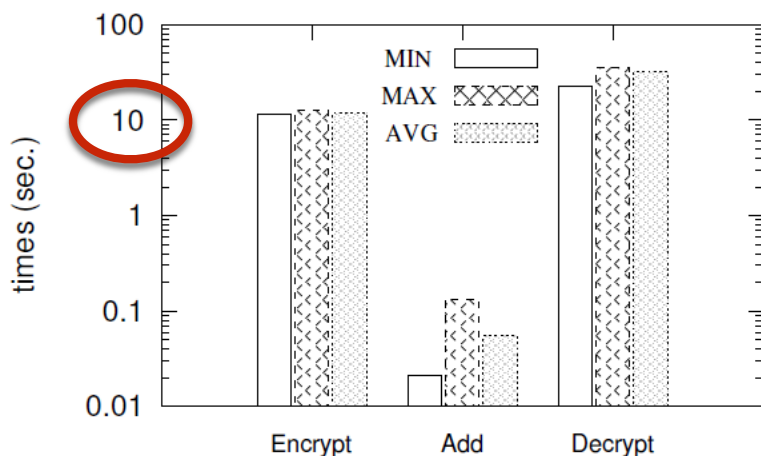
- Clustering : $k = 50$ centroids, CER dataset, 24 numbers per time-series
- Security : differential privacy budget $\epsilon = 0.69$, encryption key length 1024 bits



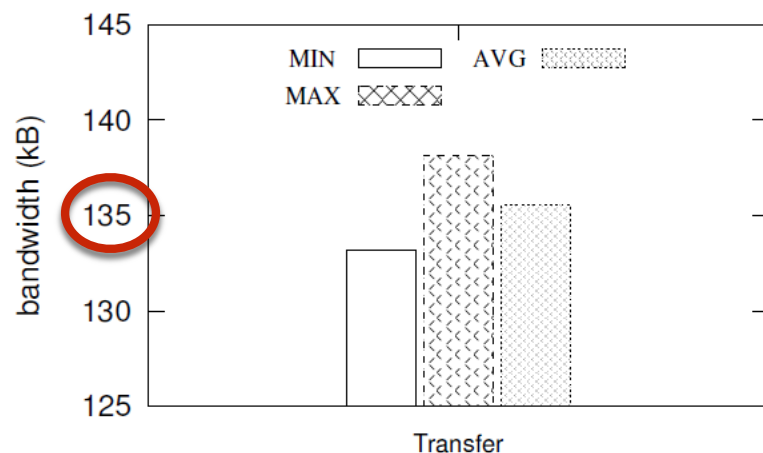
PRIVACY-PRESERVING TIME-SERIES CLUSTERING

- **Affordable communication and computation costs**

- NUMED dataset: 1.2M time series of size 20



(a) Times Consumption for Encrypting or Decrypting One set of Means, or Adding Together Two Sets of Means (seconds)



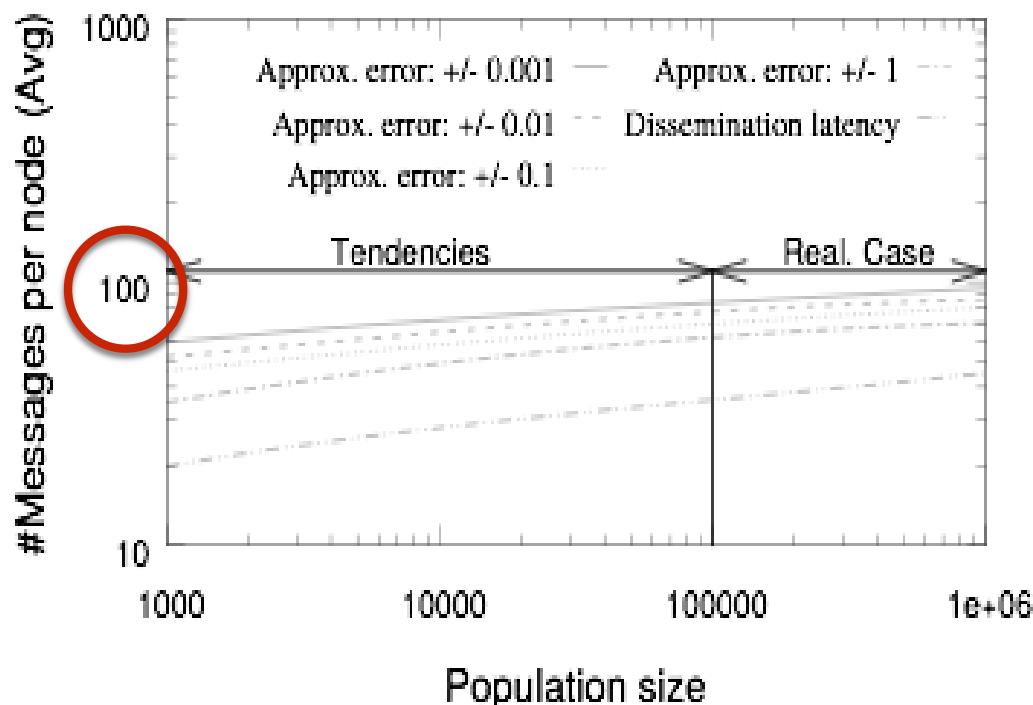
(b) Bandwidth Consumption for Transferring One Set of Means (kilo-bytes)

Unitary Local Costs for a Set of 50 Means, 20 Measures per Mean, and a 1024 Bits Encryption Key

PRIVACY-PRESERVING TIME-SERIES CLUSTERING

- **Affordable communication and computation costs**

- NUMED dataset: 1.2M time series of size 20



CONCLUSION

- **Chiaroscuro :**

- First massively distributed privacy-preserving clustering solution for time series
- Clustering: *k*-means-like algorithm (simplicity)
- Distribution: Gossip-based (scalability and fault-tolerance)
- Privacy: encryption and differential privacy

- **Future work :**

- Functional representation of time series
- Malicious participants
- Other analytical algorithms

REFERENCES

“Chiaroscuro: Transparency and Privacy for Massive Personal Time-Series Clustering”, T.Allard, G.Hébrail, F.Masseglia, E.Pacitti, Proceedings of the 2015 ACM SIGMOD.

“A New Privacy-Preserving Solution for Clustering Massively Distributed Personal Times-Series”, T.Allard, G.Hébrail, F.Masseglia, E.Pacitti, 2016 ICDE demonstration.

“Differential privacy”, C. Dwork, in ICALP, 2006, p. 1–12.

“A Generalisation, a Simplification and Some Applications of Paillier’s Probabilistic Public-Key System”, Damgaard et M. Jurik, in Proceedings of the 4th International Workshop on Practice and Theory in Public Key Cryptography: Public Key Cryptography, London, UK, UK, 2001, p. 119–136.

“Gossip-Based Computation of Aggregate Information”, D. Kempe, A. Dobra, et J. Gehrke, in FOCS, Washington, DC, USA, 2003, p. 482–491, 2003.