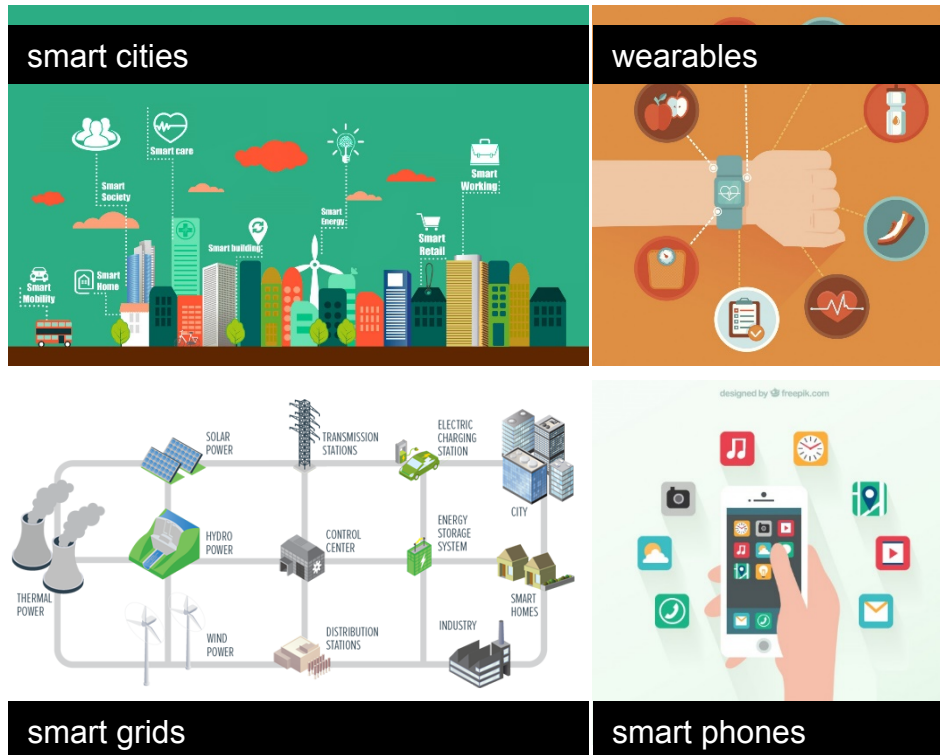




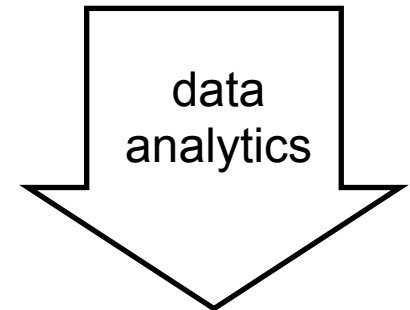
Engineering Democratization in Internet of Things Data Analytics

Evangelos Pournaras, **Jovan Nikolic**, Ales Omerzel, Dirk Helbing

Motivation



massive amounts
of unstructured
real-time data

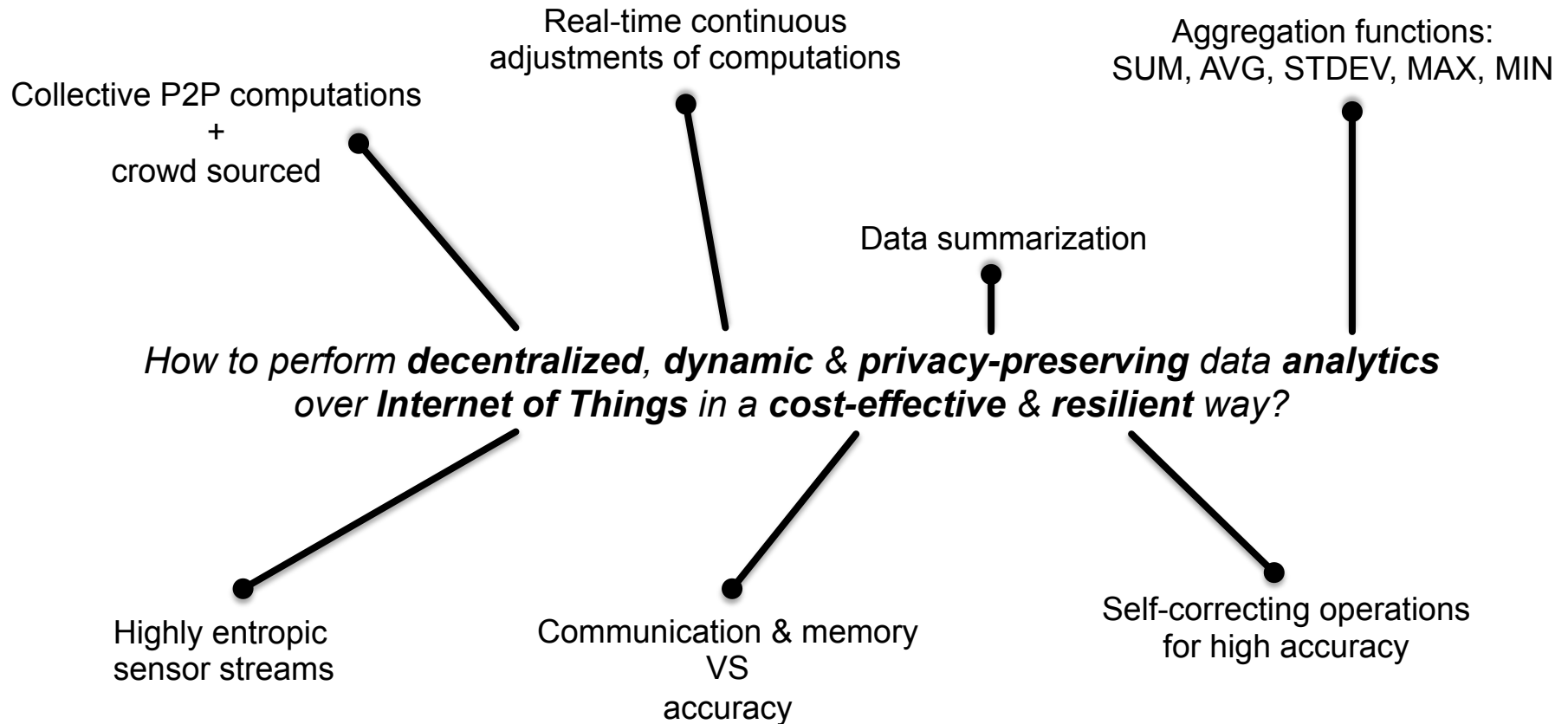


GOAL:
merchandise, service

... BUT

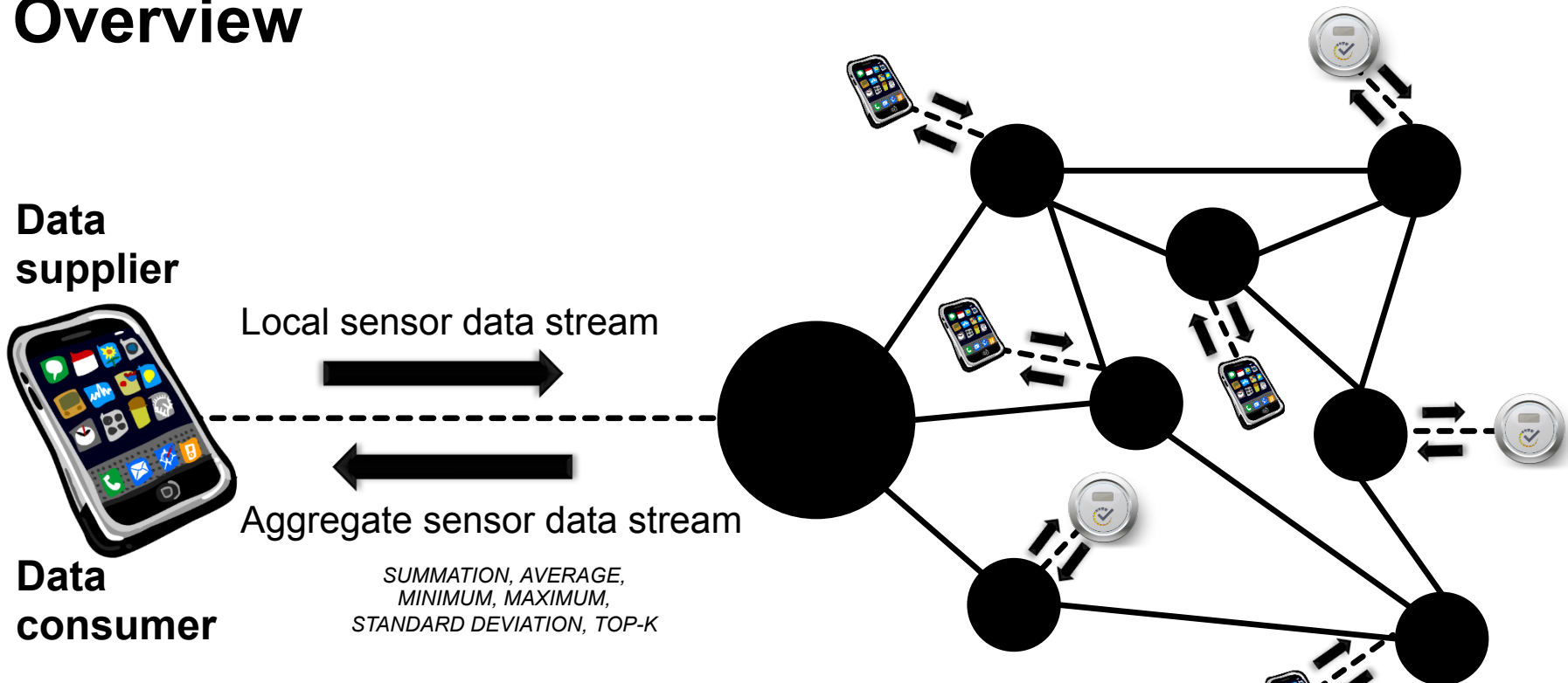
<http://younginnovator.eu/2016/09/behind-the-smart-cities-concept/>
<https://www.forbes.com/sites/reemtas/2016/10/17/the-future-of-wearables-can-companies-avoid-the-pitfalls-threatening-healthcare-wearables/#1dc9bdd16630>
<http://comsar.com/business-areas/transmission-distribution/smart-grids>
<http://www.freepik.com/free-photos-vectors/smartphone>

Research Question



DEMOCRATIZATION OF DATA ANALYTICS

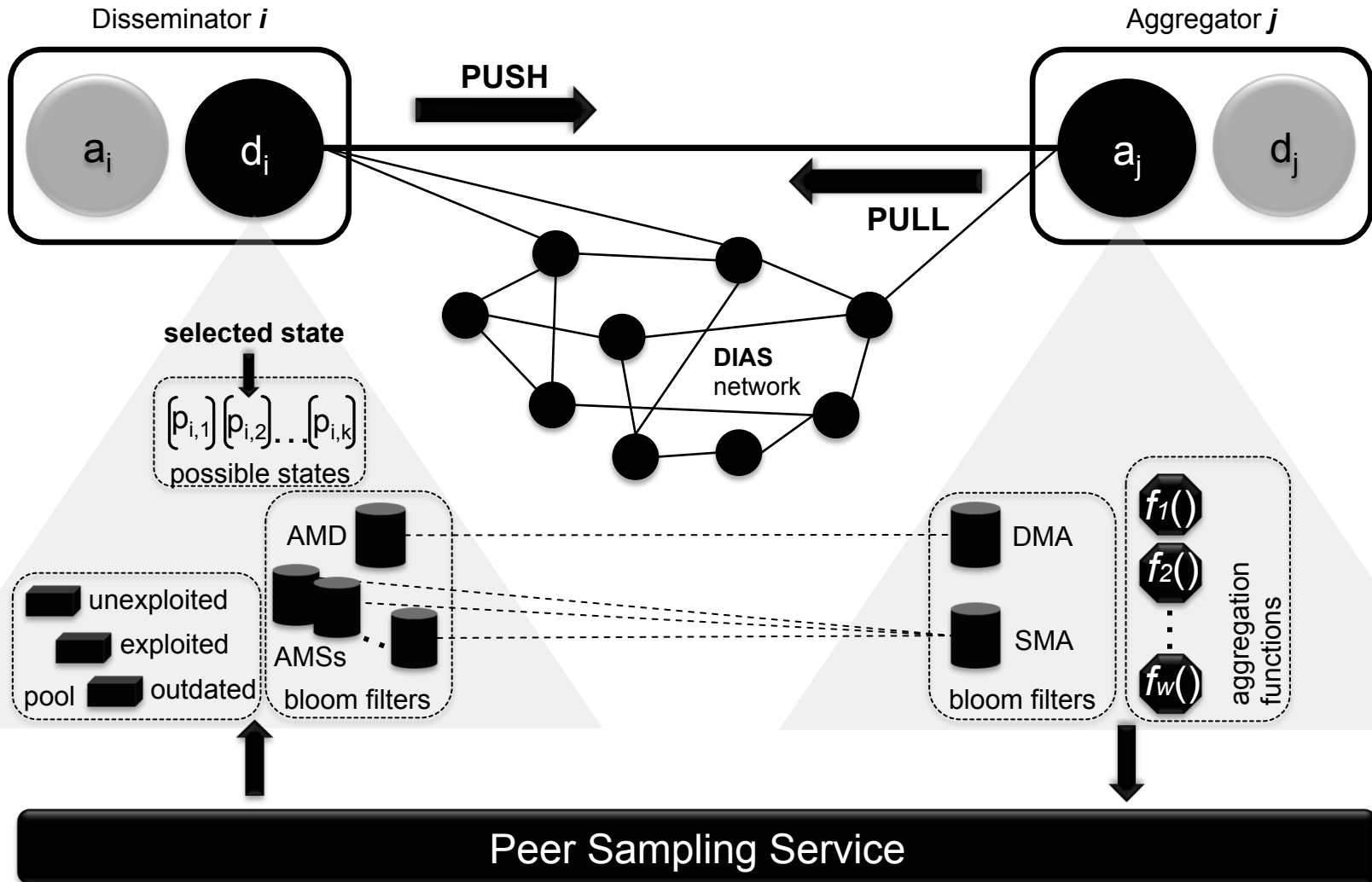
Overview



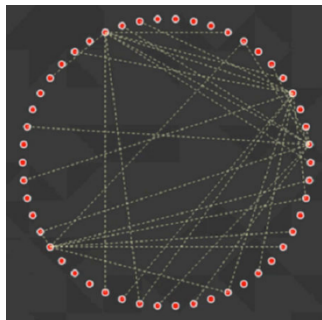
DIAS – *Dynamic Intelligent Aggregation Service:*

- local computations
- peer-to-peer interaction
- hashed information for efficiency and privacy
- self-correcting operations

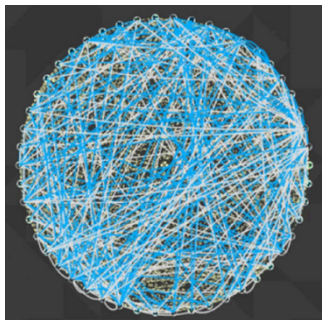
DIAS



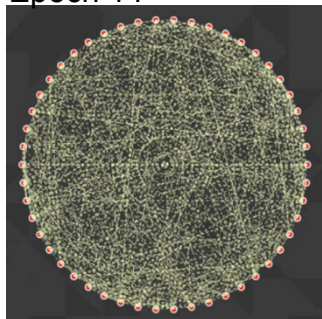
Visualization of DIAS



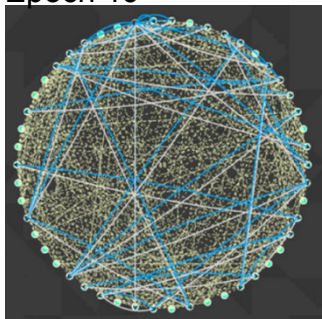
Epoch 14



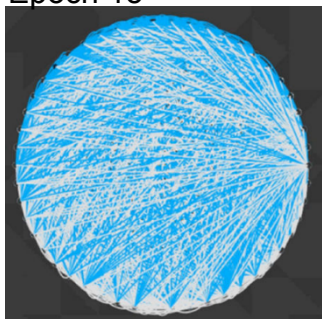
Epoch 19



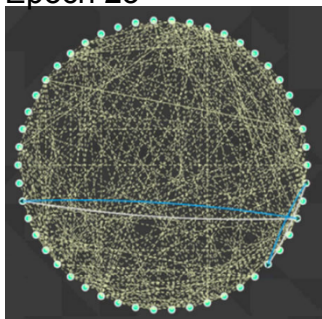
Epoch 15



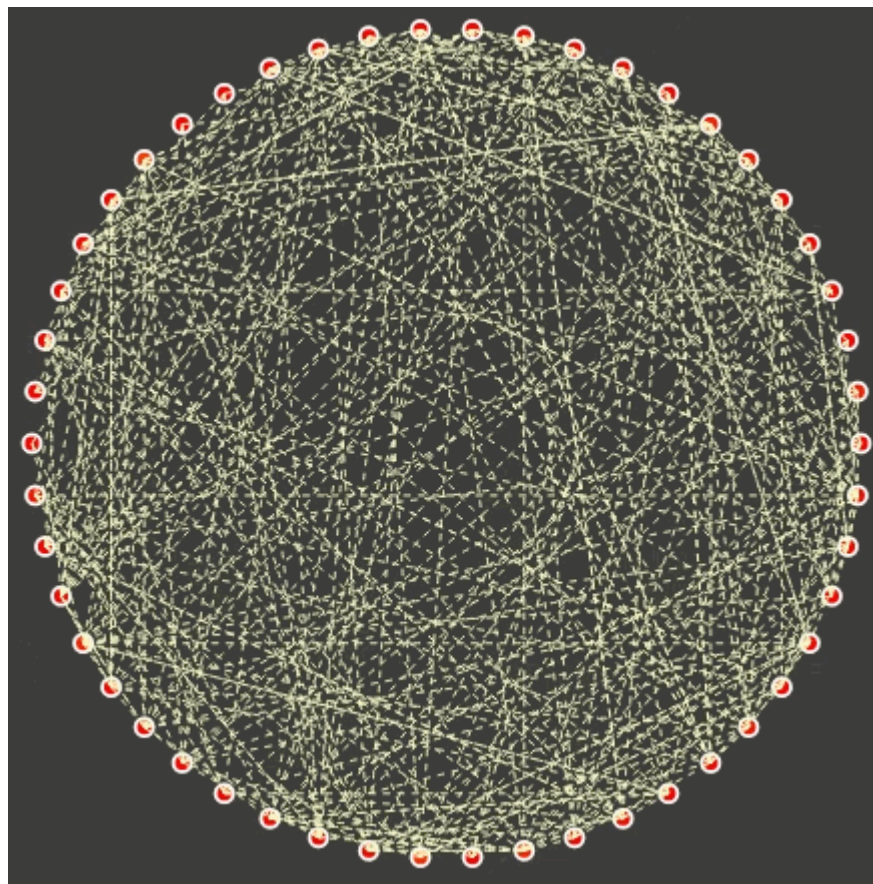
Epoch 23



Epoch 18



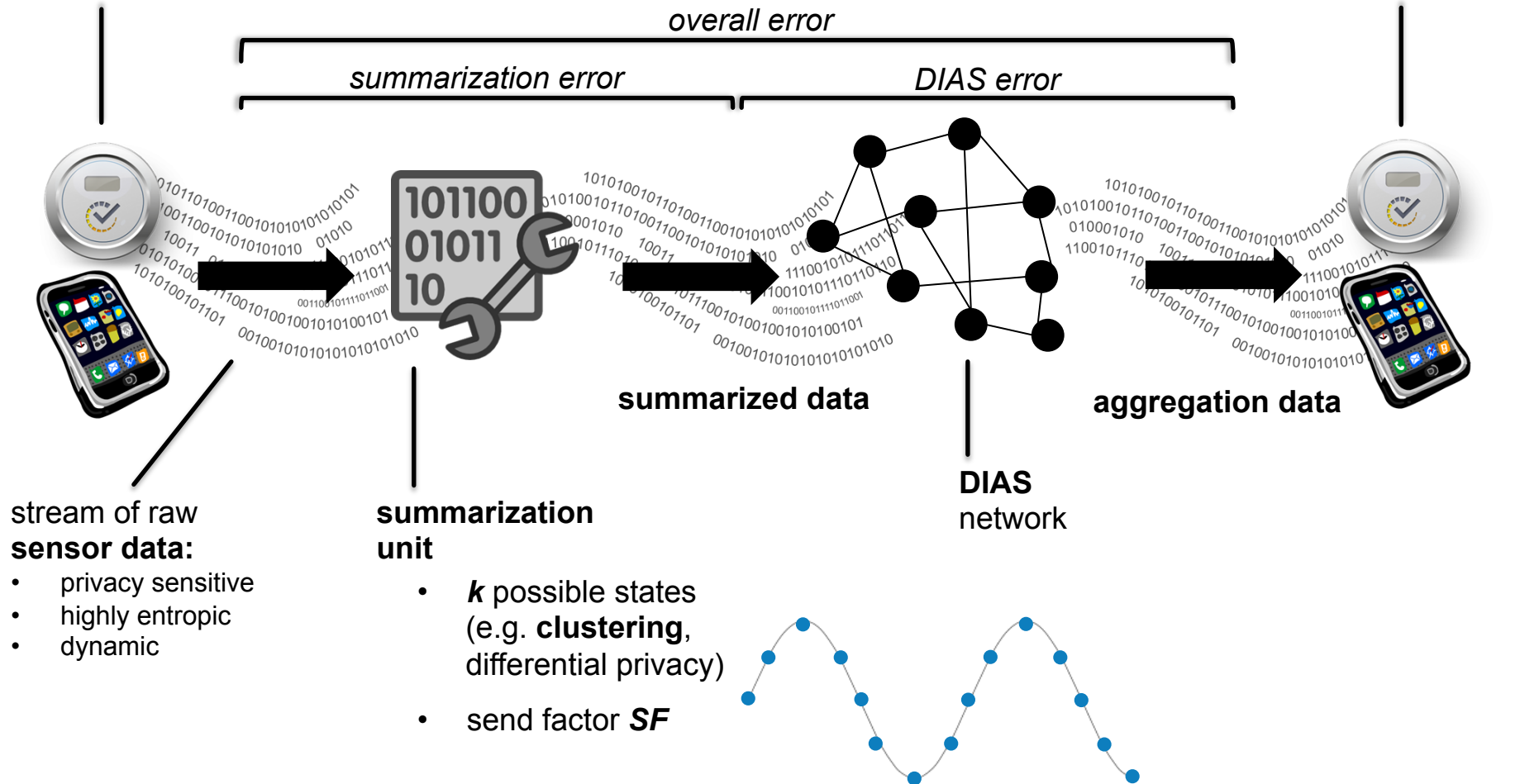
Epoch 27



Decentralized Data Management

Internet of Things
data suppliers

Internet of Things
data consumers



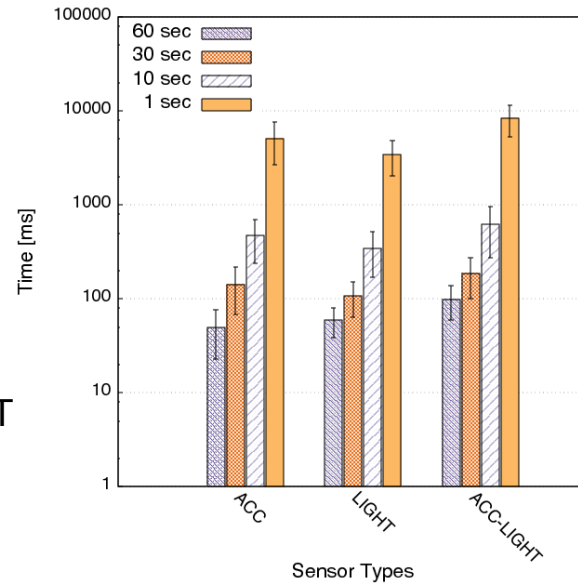
Experimental Evaluation

- 1) Computational feasibility of IoT devices
- 2) Accuracy in Analytics:
- 3) Communication costs
- 4) Trade-offs & Decision Making

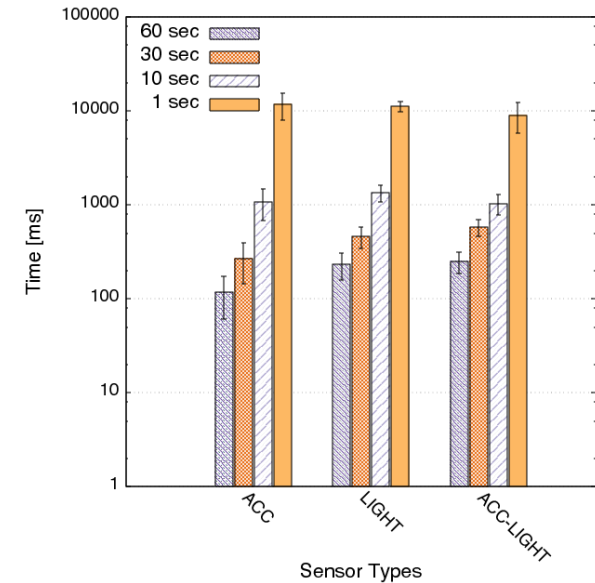
Experimental Evaluation

1. Computational feasibility of IoT devices

- 3 phones with Android OS:
 - Phone 1 (5.1)
 - Phone 2 (6.0.1)
 - Phone 3 (5.0.1)
- Sensor data:
 - ACCELEROMETER
 - LIGHT
 - ACCELEROMETER-LIGHT
- Varied frequency of sampling between 1 and 60 sec
- Average execution time over 30 repetitions



(a) Phone 1, $k = 3$

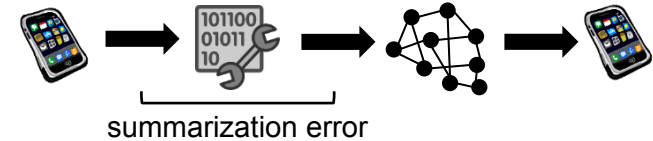


(a) Phone 1, $k = 5$

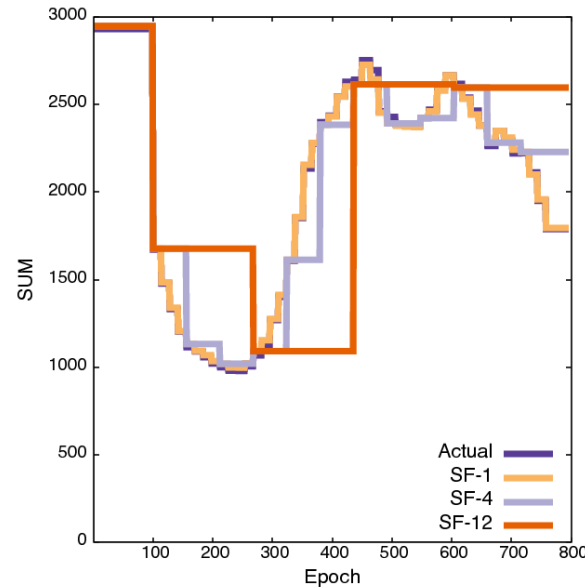
~ 50% ↑

Experimental Evaluation

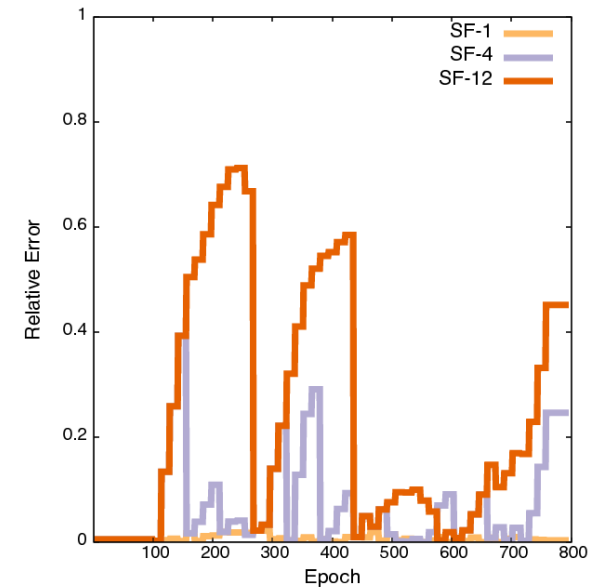
2. Accuracy in Analytics



- **ECBT** – Electricity Customer Behavioral Trial (2009-2010):
 - data from 04.01.2009.
 - collected every 30 minutes
 - total of 48 records per day
 - total of 6435 consumers
 - 3000 residential consumers
- $k = 5$ possible states
- Varied send factor:
 - **SF-1, SF-2, SF-4, SF-8, SF-12, SF-16**



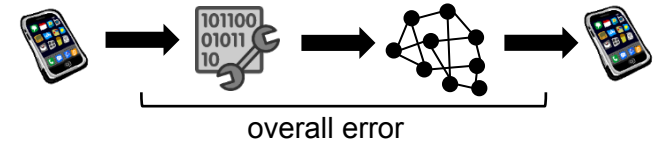
(a) summarized data



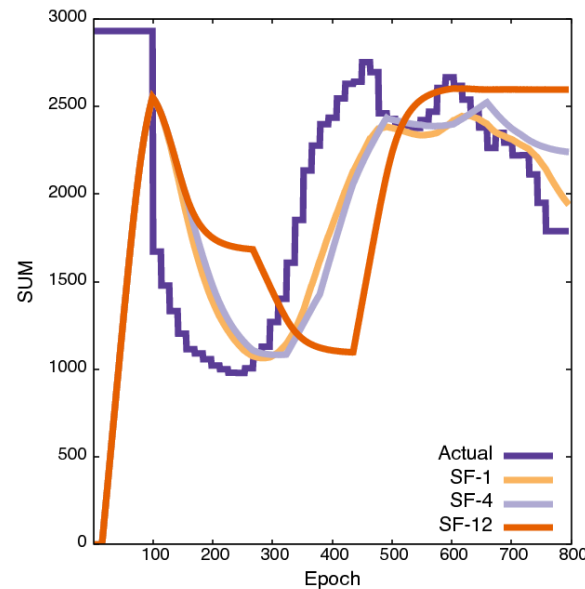
(b) summarization error

Experimental Evaluation

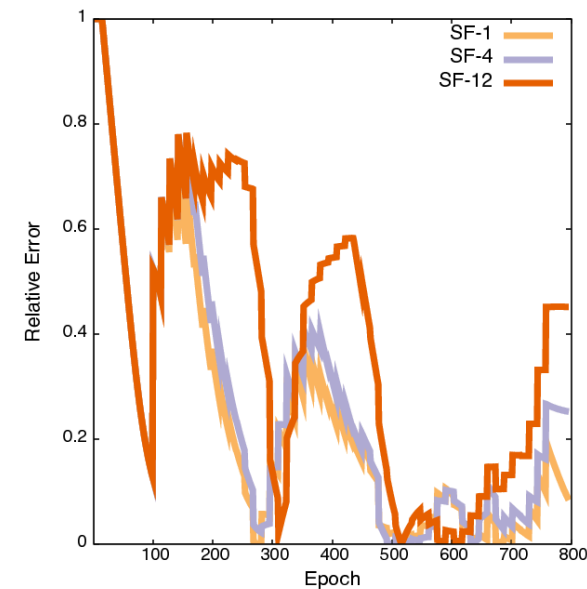
2. Accuracy in Analytics



- **ECBT** – Electricity Customer Behavioral Trial (2009-2010):
 - data from 04.01.2009.
 - collected every 30 minutes
 - total of 48 records per day
 - total of 6435 consumers
 - 3000 residential consumers
- $k = 5$ possible states
- Varied send factor:
 - **SF-1, SF-2, SF-4, SF-8, SF-12, SF-16**



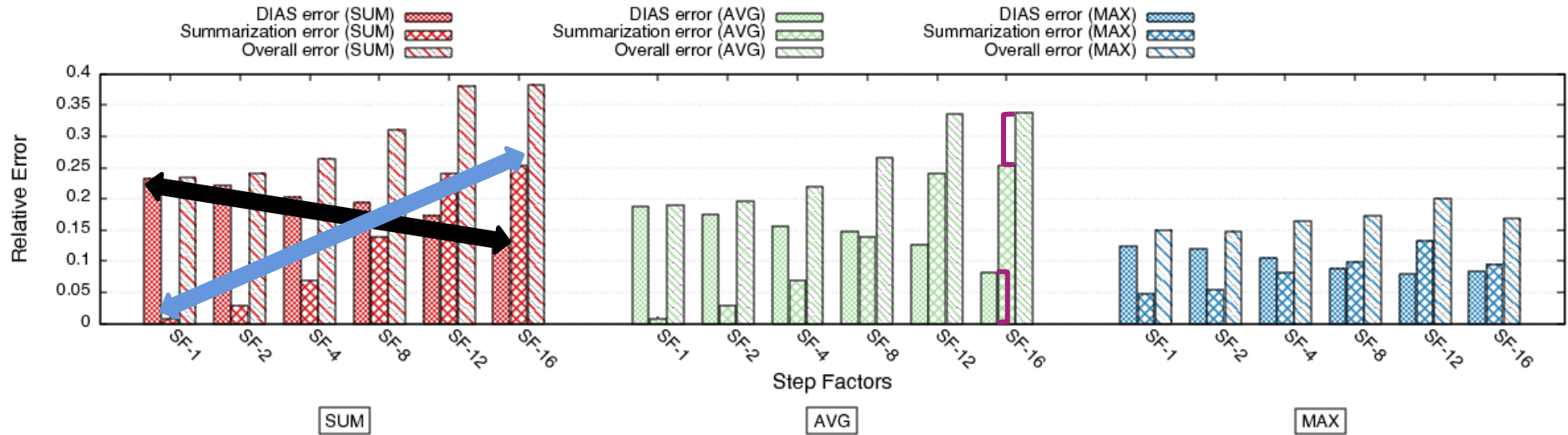
(a) aggregation data



(b) overall error

Experimental Evaluation

2. Accuracy in Analytics – Overview

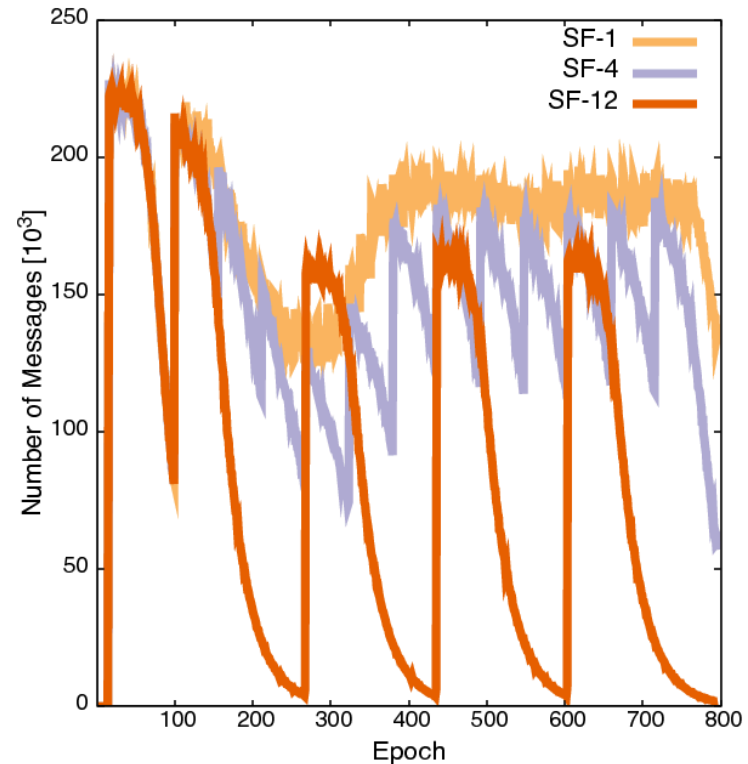


- DIAS error ↓ as send factor ↑
- summarization error ↑ as send factor ↑
- overall error (on average) is sum of its parts

Experimental Evaluation

3. Communication cost

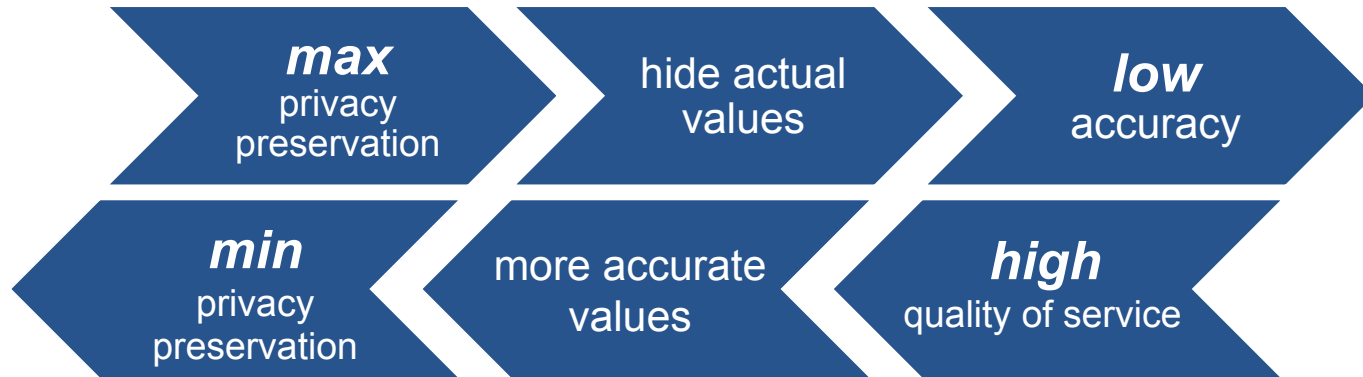
- **ECBT** – Electricity Customer Behavioral Trial (2009-2010):
 - data from 04.01.2009.
 - collected every 30 minutes
 - total of 48 records per day
 - total of 6435 consumers
 - 3000 residential consumers
- $k = 5$ possible states
- Varied send factor:
 - **SF-1, SF-2, SF-4, SF-8, SF-12, SF-16**



Number of messages sent

Experimental Evaluation

4. Trade-offs & Decision Making



Privacy regulated by:

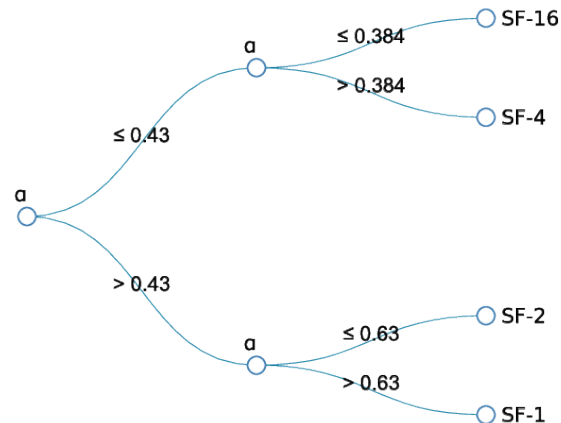
- Choice of # of possible states
- Choice of send factor

How to choose SF?

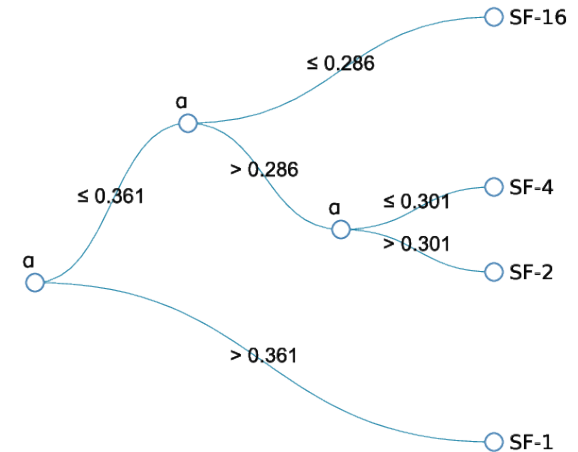
Experimental Evaluation

4. Trade-offs & Decision Making

- Decision trees (C4.5 algorithm):
- Automatic regulation



(b) Overall vs DIAS error



(a) Summarization vs DIAS error

Conclusion

- ✓ fully-decentralized
- ✓ generic data aggregation
- ✓ highly dynamic input data stream from IoT devices

FEASIBLE

privacy ↔ accuracy ↔ communication
cost

**AUTOMATICALLY
REGULATED**

PARTICIPATORY DATA ANALYTICS AS PUBLIC GOOD

Future work

- Changes and failures in the network
- Further enhancing privacy & security:
 - differential privacy
 - homomorphic encryption
- Supporting multiple sensor types

Questions?

ETH Zurich

Evangelos Pournaras

epournaras@ethz.ch

Jovan Nikolic

jovan.nikolic@gess.ethz.ch



www.sobigdata.eu



www.dias-net.org



www.asset-consumerism.eu