RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

Inria

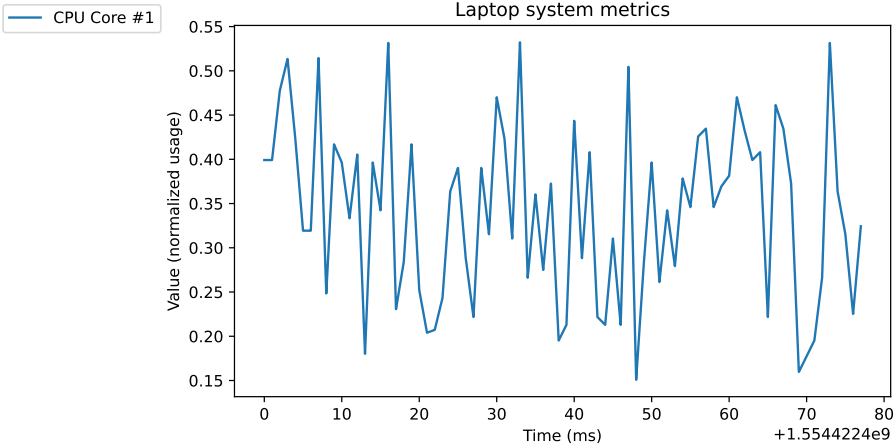
Less is More:
A Lossy Storage Middleware for Multivariate Time Series

10th GDR RSD/ASF Winter School

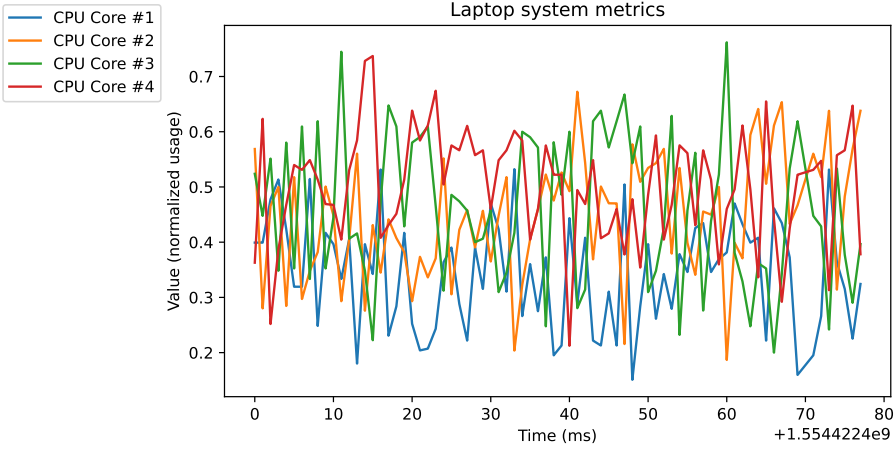
Rémy Raes

01
Context

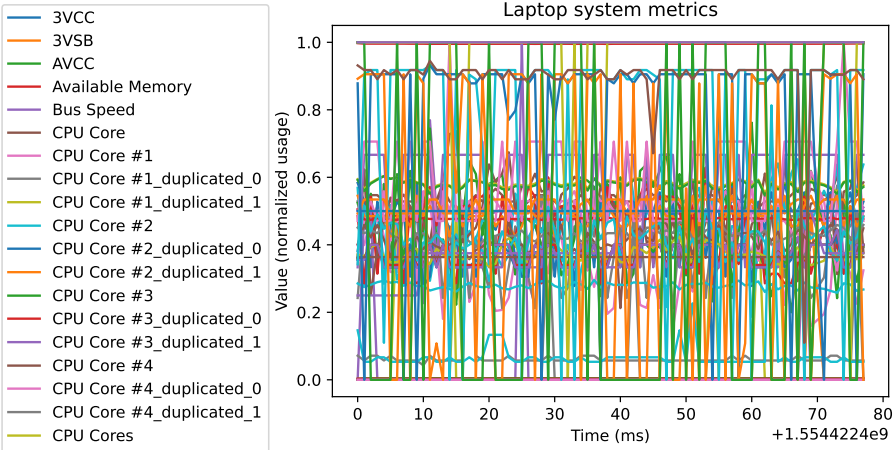
Time series


CPU Core #1
Laptop system metrics
Value (normalized usage)
0.55
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0
10
20
30
40
50
60
70
80
Time (ms)
+1.5544224e9

Time series


CPU Core #1
CPU Core #2
CPU Core #3
CPU Core #4
Laptop system metrics
Value (normalized usage)
0.2
0.3
0.4
0.5
0.6
0.7
0
10
20
30
40
50
60
70
80
Time (ms)
+1.5544224e9

High cardinality challenge



Time series storage

- Databases
  - Relational: MySQL, PostgreSQL, TimescaleDB
  - NoSQL: MongoDB, Firestore, TDengine
- Delta encoding: Gorilla
- Signal processing: FFT, Laplace transform, wavelets
- Regression (linear/polynomial)

Time series storage

- Databases
  - Relational: MySQL, PostgreSQL, TimescaleDB
  - NoSQL: MongoDB, Firestore, TDengine
- Delta encoding: Gorilla
- Signal processing: FFT, Laplace transform, wavelets
- Regression (linear/polynomial)

Time series storage

- Databases
  - Relational: MySQL, PostgreSQL, TimescaleDB
  - NoSQL: MongoDB, Firestore, TDengine
- Delta encoding: Gorilla
- Signal processing: FFT, Laplace transform, wavelets
- Regression (linear/polynomial)

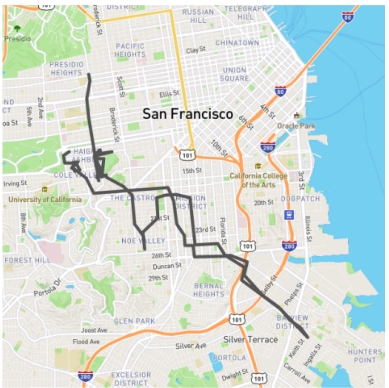
Time series storage

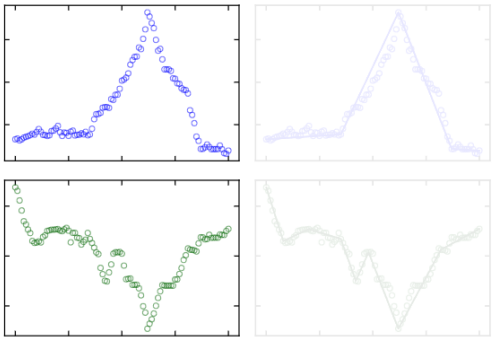
- Databases
  - Relational: MySQL, PostgreSQL, TimescaleDB
  - NoSQL: MongoDB, Firestore, TDengine
- Delta encoding: Gorilla
- Signal processing: FFT, Laplace transform, wavelets
- Regression (linear/polynomial)

Time series storage

- Databases
  - Relational: MySQL, PostgreSQL, TimescaleDB
  - NoSQL: MongoDB, Firestore, TDengine
- Delta encoding: Gorilla
- Signal processing: FFT, Laplace transform, wavelets
- Regression (linear/polynomial)

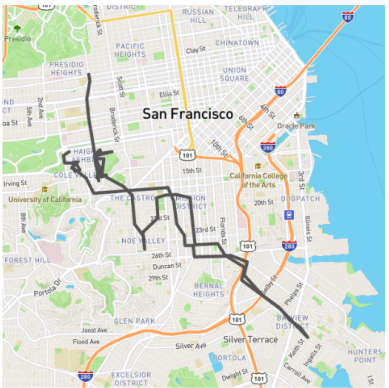
Fast linear interpolation (FLI)

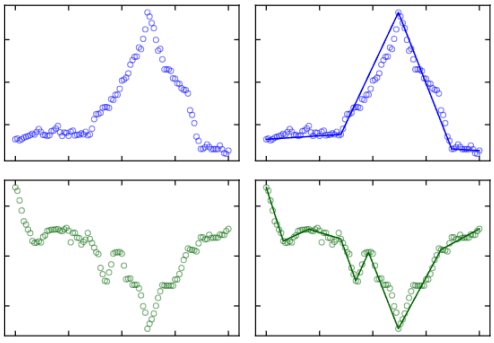




Rémy Raes, Olivier Ruas, Adrien Luxey-Bitri, Romain Rouvoy. Compact Storage of Data Streams in Mobile Devices. *DAIS'24 - 24th International Conference on Distributed Applications and Interoperable Systems, Jun 2024, Groningen, Netherlands.* **(hal-04535716)**

Fast linear interpolation (FLI)

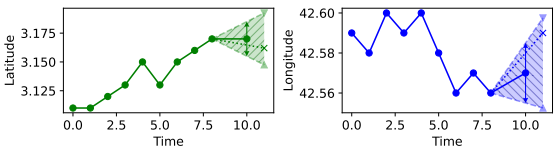




Rémy Raes, Olivier Ruas, Adrien Luxey-Bitri, Romain Rouvoy. Compact Storage of Data Streams in Mobile Devices. *DAIS'24 - 24th International Conference on Distributed Applications and Interoperable Systems, Jun 2024, Groningen, Netherlands.* **(hal-04535716)**

02
Contribution

A Lossy Storage Middleware for Multivariate Time Series



Separate modelling drawbacks

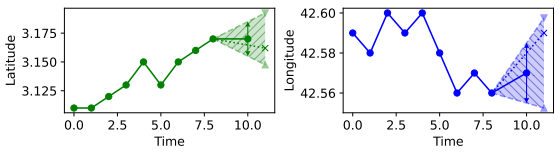
Storing series separately is not optimal, because of:

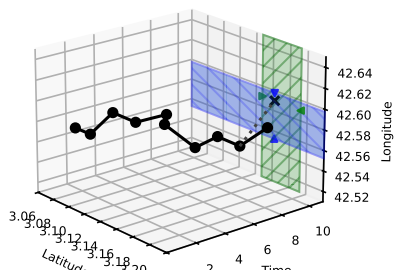
- timestamp data duplication;
- unexploited series correlation.

Combined modelling

- From Fast Linear Interpolation (FLI)...
- ...to FLInD (FLI in *n-dimension*)!

A Lossy Storage Middleware for Multivariate Time Series





Separate modelling drawbacks

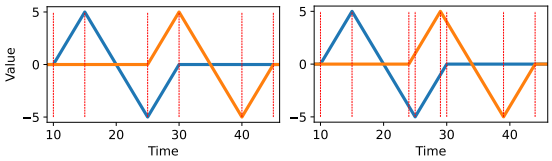
Storing series separately is not optimal, because of:

- timestamp data duplication;
- unexploited series correlation.

Combined modelling

- From Fast Linear Interpolation (FLI)...
- ...to FLInD (FLI in *n-dimension*)!

Series clustering



Shared timestamps

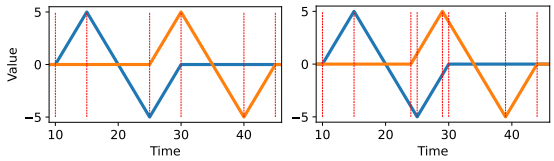
FLInD performance is directly linked to the way underlying data evolves:

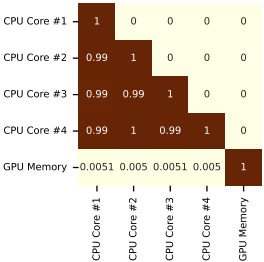
- Optimum is when all series change tendencies at the same time;
- On the other hand, time offsets undermine performances.

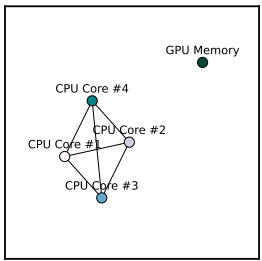
Looking for cliques

- Compute the Jaccard index of series' timestamps
- Use indexes (and a threshold value) to create a graph of series

Series clustering







Shared timestamps

FLInD performance is directly linked to the way underlying data evolves:

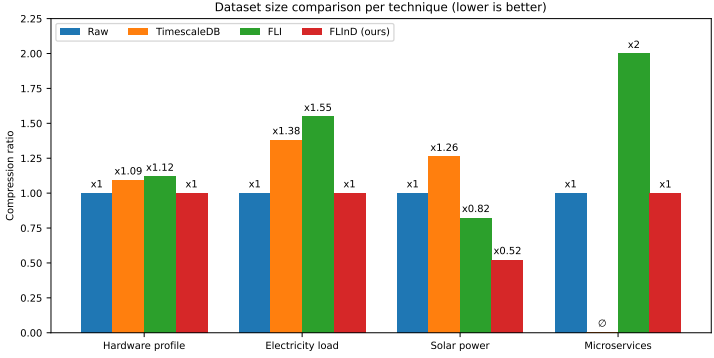
- Optimum is when all series change tendencies at the same time;
- On the other hand, time offsets undermine performances.

Looking for cliques

- Compute the Jaccard index of series' timestamps
- Use indexes (and a threshold value) to create a graph of series

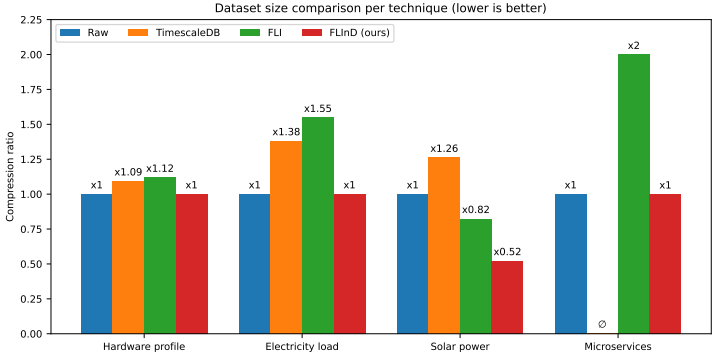
03
Results

Linear regression and multivariate series



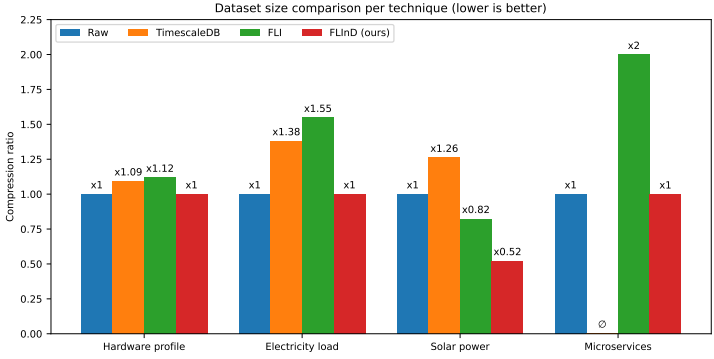
- Size gain of up to 50% with lossless compression on a few datasets;
- Same cost as raw storage in the worse case.

Linear regression and multivariate series



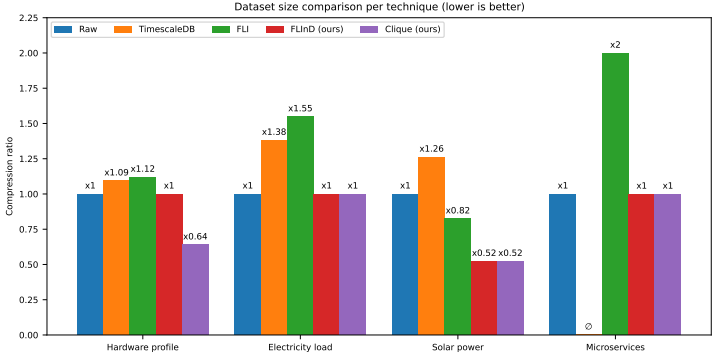
- Size gain of up to 50% with lossless compression on a few datasets;
- Same cost as raw storage in the worse case.

Linear regression and multivariate series



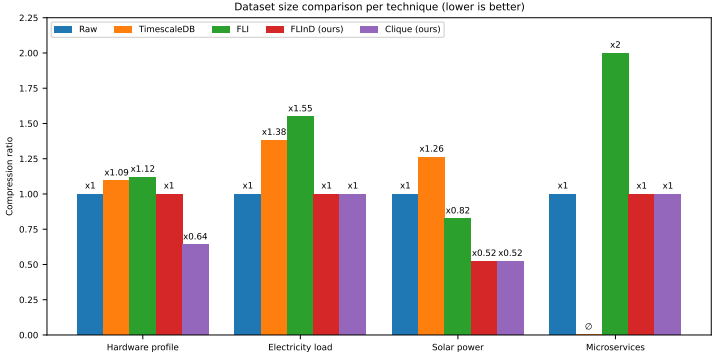
- Size gain of up to 50% with lossless compression on a few datasets;
- Same cost as raw storage in the worse case.

Series clustering



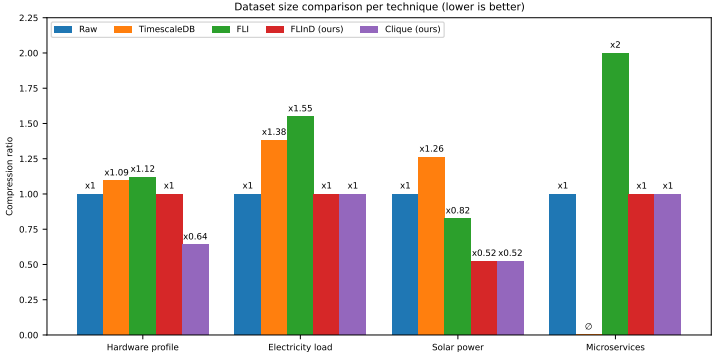
- Nice gain with the first dataset...
- ...but not elsewhere!
- Additionally, lossy compression does not work with MTS:
  - The more dimensions, the more unlikely it is all dimensions respect their invariant.

Series clustering



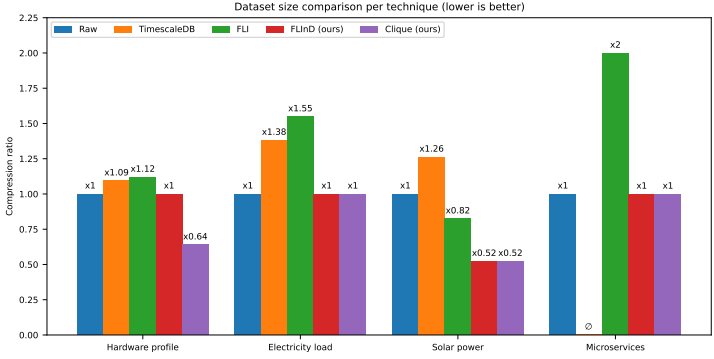
- Nice gain with the first dataset...
- ...but not elsewhere!
- Additionally, lossy compression does not work with MTS:
  - The more dimensions, the more unlikely it is all dimensions respect their invariant.

Series clustering



- Nice gain with the first dataset...
- ...but not elsewhere!
- Additionally, lossy compression does not work with MTS:
  - The more dimensions, the more unlikely it is all dimensions respect their invariant.

Series clustering



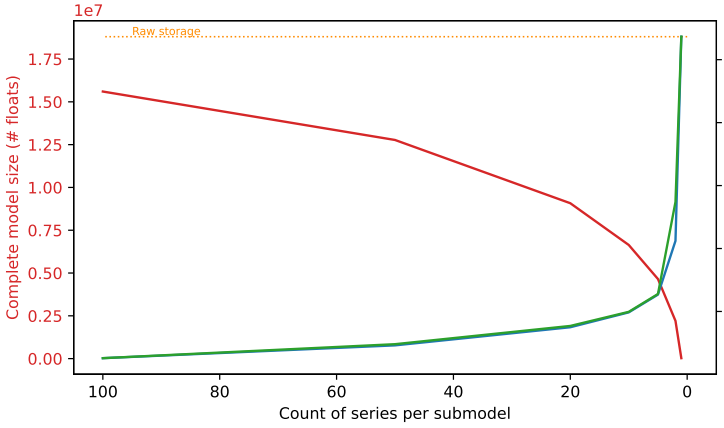
- Nice gain with the first dataset...
- ...but not elsewhere!
- Additionally, lossy compression does not work with MTS:
  - The more dimensions, the more unlikely it is all dimensions respect their invariant.

Series splitting

Name	Size	Read time (s)	Write time (s)	P95	P99	Variance
FLInD	143,27 MiB	2.634e-03	3.89e-03	0.0	0.0	1.639e-40
100 series per model	72,63 MiB	3.839e-03	5.639e-03	3.031e-02	5.467e-02	1.4e-04
50 series per model	59,76 MiB	5.527e-03	1.012e-02	3.783e-02	6.53e-02	2.026e-04
20 series per model	44,37 MiB	8.685e-03	2.093e-02	4.858e-02	7.883e-02	3.048e-04
10 series per model	35,05 MiB	1.051e-02	3.447e-02	5.705e-02	8.823e-02	3.945e-04
5 series per model	27,77 MiB	1.688e-02	5.881e-02	6.593e-02	9.698e-02	4.955e-04
2 series per model	21,27 MiB	3.433e-02	1.436e-01	7.809e-02	1.078e-01	6.505e-04
1 series per model	18,44 MiB	6.02e-02	2.70e-01	8.804e-02	1.159e-01	7.929e-04

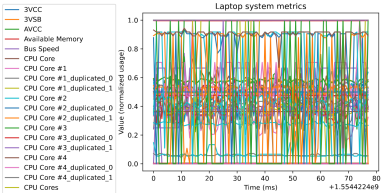
Table: Performances metrics over modelling the Microservices Traces dataset using an increasingly smaller count of dimensions per model. The lower the dimensions count, the better the error is acknowledged, leading to better total size, but also to increased read/write times due to the dataset being split between many model instances.

Series splitting

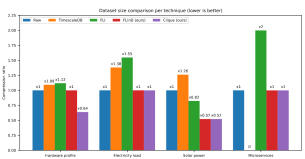


Take away

High cardinality challenge

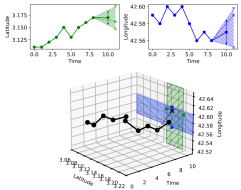


Series clustering



- Nice gain with the first dataset...
- ...but not elsewhere!
- Additionally, lossy compression does not work with MTS:
  - The more dimensions, the more unlikely it is all dimensions respect their invariant.

A Lossy Storage Middleware for Multivariate Time Series



Separate modelling drawbacks

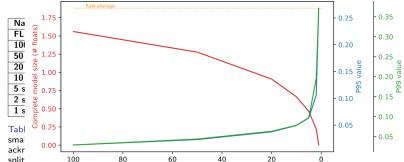
Storing series separately is not optimal, because of:

- timestamp data duplication;
- unexploited series correlation.

Combined modelling

- From Fast Linear Interpolation (FLI)...
- ...to FLInD (FLI in *n-dimension*)!

Series splitting



Na										nce
FL										e-40
10										04
50										e-04
20										e-04
10										e-04
5 s										e-04
2 s										e-04
1 s										e-04

Tabl gly
sma
ackr being
split

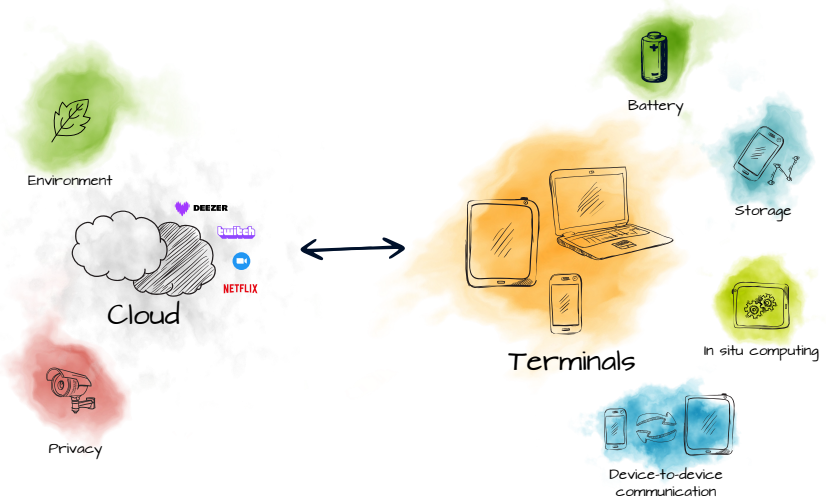
Merci.

Research questions

Distributed Machine Learning in Ubiquitous Environments using Location-dependent Models

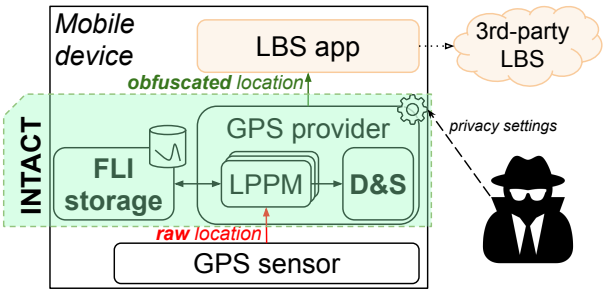
- How to store unbounded data streams on constrained mobile devices?
- How to exchange relevant model samples among nearby devices?
- How to program DML algorithms for the masses?

Ph.D overview



Embedded, mobile privacy framework

- INTACT: *in situ* location protection

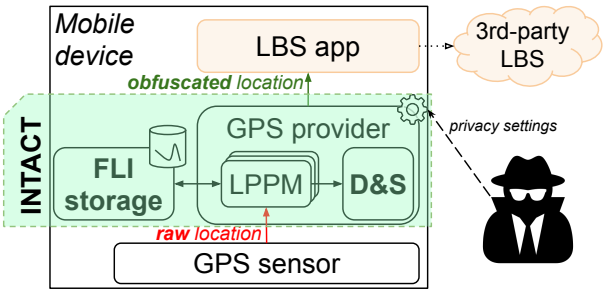


- Local storage of private data
- Local attack and protection mechanisms
- Locally ensure data is safe before sharing it

Rémy Raes, Olivier Ruas, Adrien Luxey-Bitri, Romain Rouvoy (2025).
INTACT: Compact Storage of Data Streams in Mobile Devices to Unlock User Privacy at the Edge.
Journal of Internet Services and Applications, 16(1), 372–387. **(10.5753/jisa.2025.5242)**

Embedded, mobile privacy framework

- INTACT: *in situ* location protection

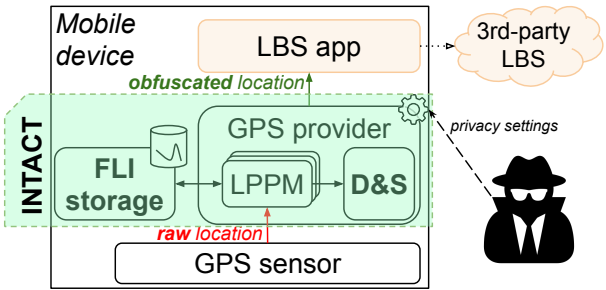


- Local storage of private data
- Local attack and protection mechanisms
- Locally ensure data is safe before sharing it

Rémy Raes, Olivier Ruas, Adrien Luxey-Bitri, Romain Rouvoy (2025).
INTACT: Compact Storage of Data Streams in Mobile Devices to Unlock User Privacy at the Edge.
Journal of Internet Services and Applications, 16(1), 372–387. **(10.5753/jisa.2025.5242)**

Embedded, mobile privacy framework

- INTACT: *in situ* location protection

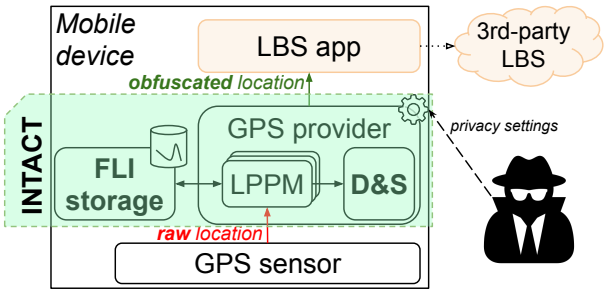


- Local storage of private data
- Local attack and protection mechanisms
- Locally ensure data is safe before sharing it

Rémy Raes, Olivier Ruas, Adrien Luxey-Bitri, Romain Rouvoy (2025).
INTACT: Compact Storage of Data Streams in Mobile Devices to Unlock User Privacy at the Edge.
Journal of Internet Services and Applications, 16(1), 372–387. **(10.5753/jisa.2025.5242)**

Embedded, mobile privacy framework

- INTACT: *in situ* location protection



- Local storage of private data
- Local attack and protection mechanisms
- Locally ensure data is safe before sharing it

Rémy Raes, Olivier Ruas, Adrien Luxey-Bitri, Romain Rouvoy (2025).
INTACT: Compact Storage of Data Streams in Mobile Devices to Unlock User Privacy at the Edge.
Journal of Internet Services and Applications, 16(1), 372–387. **(10.5753/jisa.2025.5242)**