Bern University of Applied Sciences

IARIA

# K-Area: An Efficient Approach to Approximate the Spatial Boundaries of Mobility Data with k-Anonymity

**Authors: Maël Gassmann, Annett Laube, Dominic Baumann**
**Presenter:**
**Maël Gassmann**
**Institute for Data Applications and Security**
**Email: mael.gassmann@bfh.ch**

# About me - Maël Gassmann

- Swiss
- French native speaker; German
- BSc in Computer Science in 2022 at the BFH
- MSc in Engineering $3^{rd}$ semester at the BFH
- 50% Assistant in the BFH's Institute for Data Applications and Security

# Summary

1. Introduction
2. State of the Art
3. Concept
4. Conclusion

# Introduction

- Mobility datasets
  - Complex
  - Potent in utility
- Anonymity is a **must**
- Privacy preserving methods:
  - Computationally expensive
  - Use case specific

Objectives:

- Fast assessment
- Not use case specific
- Reduce the problem space for privacy enhancing methods

# State of the Art

## Mitigation:

- Based on Heuristics
- No theoretical or provable guarantees
- Examples: *swapping*, *obfuscation*, *spacial cloaking or segmentation*

## Indistinguishability:

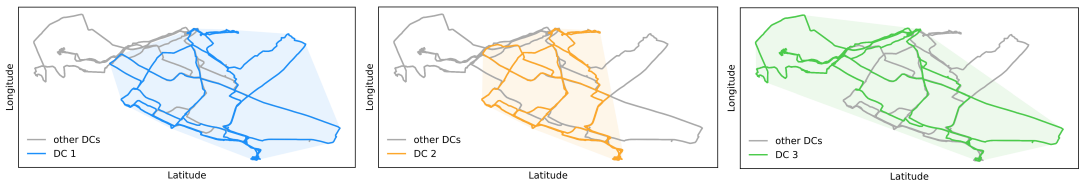
- Anonymity measured in terms of how distinguishable are each data collectors
- Filtering out singularities → better anonymity
- Examples: *k-anonymity*, ***k-area***

## Uninformativeness:

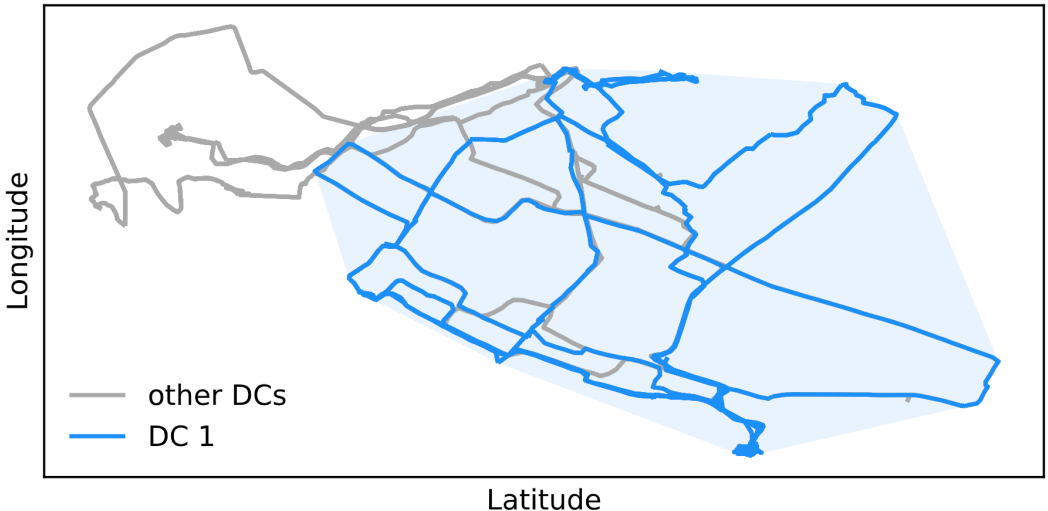
- Anonymity measured in terms of how much information is held per data collectors
- The less of difference → the better the anonymity
- Example: *differential privacy*

# Concept

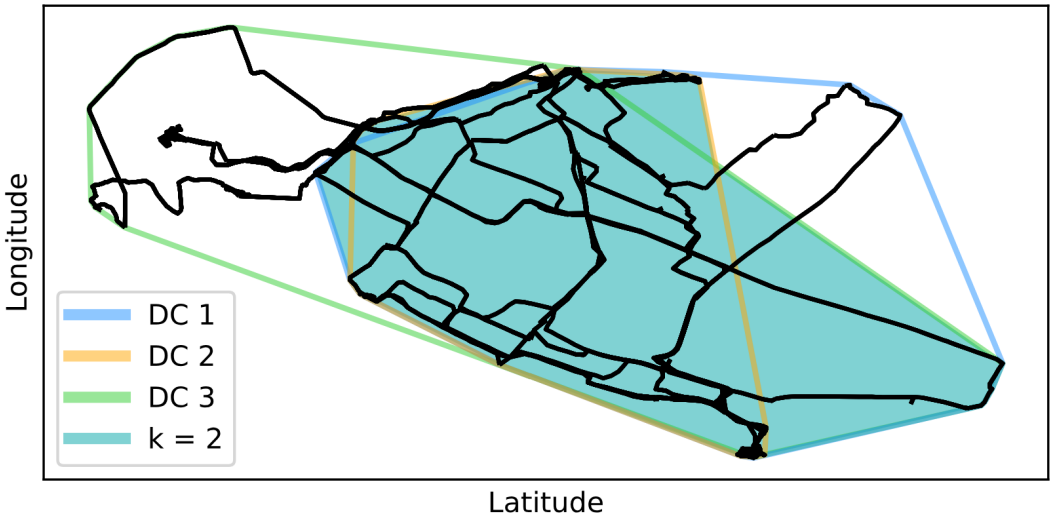
- Mobility dataset as a finite set of:
  - GPS points
  - Data collector identifiers
- Always has clear spacio-temporal bounds
- Data Collector spacio-temporal bounds
- K-Area is based on this and strives to reduce them further



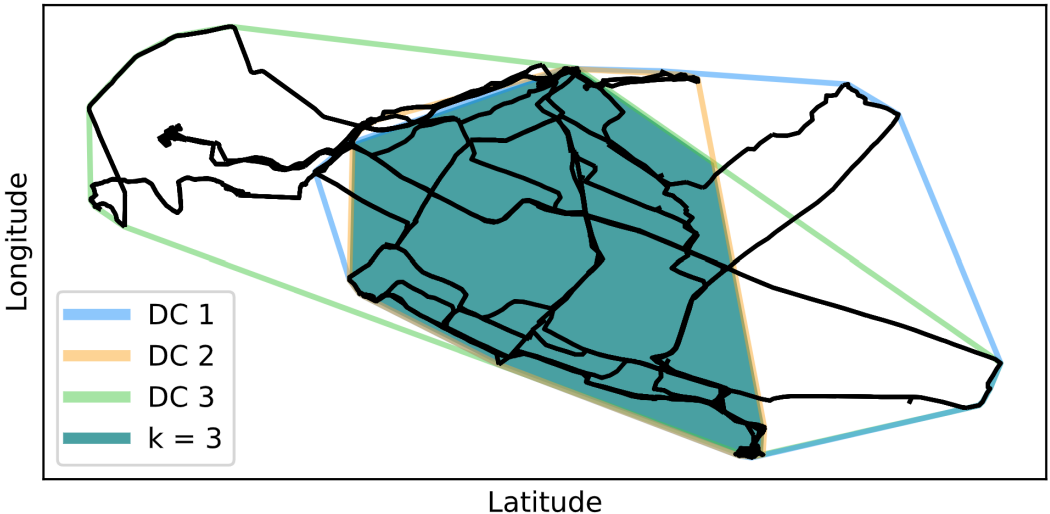
# Concept - Polygon



# Concept - K-Area



# Concept - K-Area



# Concept - Algorithm

$$A_k = \bigcup (p_{i1} \cap p_{i2} \cap \cdots \cap p_{ik})$$

for all $p_{i1}, p_{i2}, \ldots, p_{ik} \in P$

with $i_1 \neq i_2 \neq \ldots \neq i_k$ and $k \geq 2$

# Concept - Algorithm

```
1:  function K-Area(P, k)
2:      A ← array of size k with each entry as an empty set of polygons
3:      for all p_u ∈ P do
4:          for i ← k to 1 do
5:              if i > 1 then
6:                  A[i] ← A[i] ∪ (p_u ∩ A[i − 1])
7:              else
8:                  A[i] ← A[i] ∪ p_u                          ▷ i = 1
9:              end if
10:         end for
11:     end for
12:     return A
13: end function
```

# Concept - Use cases

- Set a k-anonymity condition and run the algorithm periodically while collecting data. Allowing for better management of surveys during their production.
- Generate heat-maps to visualize the readiness of a dataset and where data could be lacking.
- To be used as a pre-processing step before running computational expensive algorithms.

# Conclusion

- A solution was proposed:
  - To quickly assess a dataset
  - That can be used to reduce the problem space of privacy enhancing methods
  - Is not use case specific
  - Discards superfluous parts
- Based on geometric operations
- Polynomial order

Future work:

- Different shapes for the spacial bounds
- Researches concerning its application in 3 dimensions

# Questions

Questions ?